Modelling and empirical evaluation of labour supply behaviour

Woittiez, Isolde Beatrijs

Publication date:
1990

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Woittiez, I. B. (1990). *Modelling and empirical evaluation of labour supply behaviour: Emphasis on preference formation, job characteristics and hours restrictions*. [Doctoral Thesis, Tilburg University]. [s.n.].

MODELLING AND EMPIRICAL EVALUATION OF LABOUR SUPPLY BEHAVIOUR

emphasis on preference formation, job characteristics and hours restrictions

ISOLDE WOITTIEZ

MODELLING AND EMPIRICAL EVALUATION OF LABOUR SUPPLY BEHAVIOUR

emphasis on preference formation, job characteristics
and hours restrictions

Proefschrift ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. R.A. de Moor, in het openbaar te verdedigen ten overstaan van een door het college van dekanen aangewezen commissie in de aula van de Universiteit op

vrijdag 11 mei 1990 te 14.15 uur

door

ISOLDE BEATRIJS WOITTIEZ

geboren te Rotterdam

Promotor: Prof. Dr. Ir. A. Kapteyn

Acknowledgements

The research presented in this thesis was carried out at the Department of Econometrics of Tilburg University. I finished writing the book while working at the Department of Economics of the Faculty of Law at The University of Leiden. Data were provided by the Netherlands Central Bureau of Statistics (CBS) and the Organization for Strategic Labour Market Research (OSA).

I am very grateful to my promotor Arie Kapteyn for his enthusiastic and instructive guidance during the research period. Of the many colleagues in Tilburg and Leiden who have been helpful with my thesis in one way or the other, I would like to mention a few in particular. I am much indebted to Arthur van Soest for all his inspiring and constructive comments, and to Martijn Tummers who co-authored the paper on which Chapter 6 is based. Furthermore, I would like to thank Marcel Kerkhofs, Maarten Lindeboom and Gusta Renes for conscientiously proof-reading the entire manuscript and for their helpful comments.

Special thanks go to my parents for their unwavering practical and moral support, their constant encouragement and their sincere interest in what I did while working on my thesis and long before that in all the preceding stages.

Finally, I am particularly grateful to Frank for remaining cheerful while I spent my leisure working on this thesis. I very much appreciate both his moral contribution in the sense of interest and stimulation, and his actual contribution in the form of drawing the figures and designing the cover.

Isolde Woittiez
Rotterdam, March 19, 1990

Contents

1 Introduction

One of the major issues of policy makers in The Netherlands is to reduce the high unemployment rate. In 1988 economic growth was substantial in all OECD countries, which led to an increase in employment. The economic growth also induced extra labour supply, especially of married women, which altogether led to a smaller reduction in the unemployment rate than could have been expected in view of the economic growth (see Rapportage Arbeidsmarkt, 1989). The estimated official unemployment rate in 1988 is still 11.0% of the total labour force. Therefore, there is a strong interest in policies that seek to increase employment (the demand side of the labour market) as well as in understanding the factors that influence labour supply. In this thesis we try to further such understanding by constructing a detailed model of household labour supply.

The data we use relate to Dutch households in 1985. In that year the official rate of unemployment was 15.9%. A distinguishing feature of Dutch labour supply is its very low level of female labour force participation, e.g. in 1985 it was only 35.2%. Apart from Spain, which had a similar participation rate, most other industrialized OECD countries had a participation rate of around 60% (see OECD Labor Force Statistics). As there seems to be no obvious reason why in the long run female labour force participation in The Netherlands should remain systematically below that in other European countries, women could potentially cause a substantial increase in total labour supply.

In this thesis labour supply behaviour of individuals is analysed within a neoclassical framework of utility maximization. Several groups of individuals are distinguished: single males, single females, married males , and married females. For each of these groups we have estimated different models that are developed in the thesis. In the final chapter we will compare the different models with respect to their implied wage elasticities of labour supply for these groups.

In Chapter 2 we present the model that serves as a starting point for the other models, developed in the subsequent chapters. We analyse a household consisting of one individual who is able to work, and who is faced with the choice on how much to buy of a composite consumption good at a given price and on how much time to spend on leisure (and not work).

The number of hours the individual prefers to work and the quantity of the consumption good he desires follow from a standard maximization problem in which utility is maximized subject to a linear budget constraint. To estimate such models requires the specification of a particular functional form for the utility or demand functions. Throughout the larger part of this thesis we have chosen the same specification for the utility function. This so-called standard model consists of an hours equation that is quadratic in wages and of a consumption equation. The influence of family composition is modelled by making one of the parameters dependent on the number of family members. Closely related to this standard model is a two-adult household model, in which the joint labour supply decision of husband and wife is analysed.

An extensive description of the data can also be found in Chapter 2. Most models developed in the thesis have been estimated on two data sets, the OSA-survey and the SEP. OSA stands for "Organization of Strategic Labour Market Research" and SEP for "Social Economic Panel". The OSA-survey was held in 1985 and only persons between 18 and 60 were interviewed. The SEP is a bi-annual panel of households who are interviewed in April and October of each year, starting from 1984. For our analysis we used the October 1985 wave to make it comparable with the OSA survey. Information on both working and nonworking individuals is used in a Tobit-type likelihood function. Since wage rates are not observed for nonworking individuals a wage equation in terms of age and education has been estimated on working individuals allowing for possible selection bias. The wage predictions from this equation are inserted in the the labour supply models in Chapters 2 through 4. In Chapters 5 and 6 simultaneous hours-wage models are estimated.

Chapters 3 through 6 consist of a fair number of extensions of the standard model, described in Chapter 2. In Chapter 3 we abandon the assumption of a convex budget constraint. The Netherlands has a complex social security and welfare system and a highly nonlinear progressive tax system. We concentrate on the social security and welfare system. If an unemployed person receives unemployment benefits and plans to accept a job, he or she will loose these benefits. This introduces nonconvexities in the budget constraint and complicates estimation as nonconvexities require utility comparison.

Another adjustment made in Chapter 3 is the use of preferred hours rather than actual hours. Preferred hours is the number of hours an individual prefers to work instead of the number of hours he or she actually works. Since many individuals are not completely free in choosing the hours they work, actual and preferred hours do not always coincide. By using preferred hours we have a model of labour supply in which demand side restrictions do not play a role. In Chapter 6 we will develop a model in which demand side constraints are explicitly taken into account.

Chapter 4 emphasizes the differences in utility functions across individuals due to preference formation. Preference formation refers to the phenomenon that someone's present preferences depend on his own past behaviour (habit formation) and on other individuals' (past) behaviour (preference interdependence). Habit formation is modelled by making one of the parameters in the model, developed in Chapter 3, dependent on hours worked by that individual in the previous year. Preference interdependence is incorporated by making that same parameter also dependent on the mean hours worked by individuals in one's social reference group. In Chapter 4 we discuss two methods to construct social reference group variables. In the first method the reference group variables are constructed on the basis of a set of assumptions. In the second method we exploit the direct information on reference groups in the SEP-survey, to construct a number of indicators for the mean hours worked and mean family size in the reference group. With the help of a factor analysis model we relate the "true" and unobeserved reference group means to the observed indicators.

In Chapters 2 through 4 the utility function has hours of work and consumption as arguments. In Chapter 5 we extend the utility function with job amenity as an extra argument. The job amenities are for example social status and good working conditions. Including these allows one to model the labour supply decision and job choice simultaneously. We adopt the home production approach to model this joint decision. (See e.g. Pollak and Wachter (1975).) In the same chapter we examine two different specifications for the production function of the job amenity. In a first specification both job characteristics and hours worked are part of the job amenity. Loosely speaking, this specification allows for the fact that working many hours in a tedious job is worse than working only a few hours in that same job. In the second specification the amenity of a job is independent of the hours worked in that job. One could think of status of a job as a plausible interpretation in this case.

The job choice involves a trade-off between the wage rate and the job amenity, the idea being that the rewards for working can be both pecuniary (money wage) and nonpecuniary (desirable job characteristics). As a result the budget constraint is nonlinear.

In Chapter 6 we take a first step towards integration of labour supply and labour demand into one model. These models are especially important if labour demand and labour supply are not well matched, which is presumably the case for married women looking for part time jobs, while part time jobs are rarely offered. Labour demand restrictions are incorporated by assuming that employers offer jobs with a fixed number of hours. Workers face the market distribution of these employment opportunities. It is furthermore assumed that the market distribution of job offers is the same for all individuals. An individual chooses from the available job offers the one that yields highest utility. Note that in this model the individual can no longer <u>freely</u> choose the number of hours he prefers to work. Another extension in this chapter is that the before tax wage rate is made dependent on hours of work. Rosen (1976) justifies this procedure by suggesting that there might exist different markets for jobs with varying numbers of hours.

To avoid a forbiddingly high degree of complexity of this model, we have chosen for a somewhat simpler specification of the utility function, i.e. one with a linear rather than a quadratic labour supply function.

Chapter 7 provides an overall evaluation of the estimation results. In the first part of Chapter 7 we test the hypothesis that both samples used for estimation have been drawn from the same subset of the Dutch population. In the second part we compare labour supply wage elasticities, generated by the different models and samples. Likelihood ratio and goodness of fit statistics are presented as criteria for choosing between alternative models.

Finally, in Chapter 8 we summarize and propose issues for further research.

2 A simple neoclassical model of labour supply

2.1 Introduction

In this chapter we point out how labour supply behaviour can be analysed within the framework of utility maximization subject to a linear budget constraint. The model we use is neoclassical in the sense that the individual is assumed to maximize utility subject to time and price constraints. Quite a few researchers have taken the neoclassical model as a starting point of their analysis of labour supply (cf. Pencavel (1986) and Killingsworth and Heckman (1986)).

In Section 2.2 and 2.3 a neoclassical model is extensively described that is used as a starting point of all models developed in the following chapters. We first consider households with only one adult (single adults or adults with children) in Section 2.2. Attention is paid to households with at least two adults in Section 2.3. All models developed in this thesis have been estimated on (either of) two different data sets, the OSA-survey and the SEP. In Section 2.4 we describe both data sets, and in Section 2.5 estimation results are presented. Section 2.6 concludes.

2.2 The model specification for one-adult households

The labour supply of the adult is assumed to be consistent with the following model

$$h_k = \delta + \gamma w_k + \beta[I_k + \vartheta + \delta w_k + 1/2\gamma w_k^2] \qquad (2.1)$$

where h_k = number of hours the individual k would like to work per week
w_k = after tax marginal wage rate of individual k
I_k = (weekly) nonlabour income of household k
ϑ, δ, γ, β are parameters.

Equation (2.1) states that labour supply is a quadratic function of wages. The obvious advantage of a quadratic specification over a linear one (which is used for instance by Hausman (1980,1985), Blomquist (1983) or

Bekkering et al. (1987) is that it is more flexible. For instance, for certain levels of the wage rate labour supply may be forward bending (an increasing wage rate implies increasing labour supply) whereas for other wage levels it may be backward bending (an increasing wage rate implies a decreasing labour supply). With a linear specification, labour supply is either everywhere forward bending or everywhere backward bending. Another argument in favour of this specification is the relative ease with which shadow wages can be computed. For instance since we allow for nonconvexities in the budget set in Chapter 3, we have to compare utility in different points of the choice set. To be able to do this, without recourse to a direct utility function, shadow wages and shadow income are needed. For other flexible forms, such as the indirect translog and the AIDS, the computation of shadow wages is cumbersome.

Under mild conditions on the parameter values, equation (2.1) can be obtained as the result of maximization of a household utility function under a linear budget constraint. To be more exact, let the budget constraint for individual k be

$$y_k = w_k h_k + I_k \tag{2.2}$$

where y_k is total household consumption in a given period. Equation (2.2) simply says that the amount of money available for consumption is the sum of nonlabour income (e.g. from returns on investments or money earned by children whose labour supply is taken exogenous) and labour income (the number of hours worked by individual k per period times the wage rate). Next, consider the following utility function

$$U(h_k, y_k) = I_k^* \exp[\frac{\beta}{\gamma}(h_k - \delta - \beta I_k^*)] \tag{2.3}$$

where $$I_k^* = \frac{\gamma}{\beta^2} - [\frac{\gamma^2}{\beta^4} + \frac{\gamma}{\beta^2}\{(h_k - \delta)^2 \frac{1}{\gamma} - 2(y_k + \theta)\}]^{1/2} \tag{2.4}$$

Equation (2.3) says that household k derives utility from consumption and from hours worked on a paid job. The marginal utility derived from working in the market may very well be negative; in fact we could as well write (2.3) as a utility function of leisure T - h_k (where T is the total amount of time available per period) and y_k. On the right hand side of (2.3) we would then everywhere replace h_k by $T - L_k$ with L_k the amount of leisure enjoyed by the adult in household k.

Under a mild regularity condition (Kapteyn, Kooreman, Van Soest (1989)), one can straightforwardly show that maximization of $U(h_k, y_k)$ with respect to h_k and y_k under the budget constraint (2.2) yields relation (2.1) as a solution for h_k. The solution for y_k (not given here) can easily be obtained by substitution of (2.1) into (2.2). Like other researchers we have treated the qualitative implications of the theory as maintained hypotheses (cf. Burtless and Hausman (1978) and Van Soest, Kooreman and Kapteyn (1988)). In the latter paper it is noted that if in more complex models such as those dealing with nonlinear budget constraints or corner solutions, the Slutsky restrictions are violated, the endogenous variable may no longer be determined unambiguously by the model. For that reason appropriate restrictions on the parameters have to be imposed. For the specification that we use only concavity of the cost function has to be imposed a priori, since homogeneity and monotonicity with respect to u are satisfied automatically. This implies that the implications of the theory of utility maximization for labour supply analyses cannot be modelled as testable hypotheses. The concavity-condition in this model is $\gamma > \beta^2 I_k^*$, which guarantees convexity of indifference curves. It turns out that for the data described in Sections 2.4 and 2.5 the condition is satisfied at most data points. We also impose the assumption that leisure is a normal good ($\beta<0$) and that the largest part of the wage effect is positive ($\gamma>0$).

Thus, if equation (2.1) turns out to give a good description of the relation between labour supply, wages and nonlabour income, then we can interpret household behaviour as the result of maximizing the utility function $U(h_k, y_k)$ subject to the budget constraint. Of course if we estimate the parameters δ, γ, ϑ and β on the basis of empirical data, the household utility function is estimated implicitly.

The exposition so far implicitly assumed that different individuals have identical utility functions, so that variation in observed behaviour can be ascribed solely to variation in the budget constraints faced by different individuals. One can relax this assumption in various ways. One possibility is to assume only that certain homogeneous groups of individuals have the same utility function. Then one can apply the procedure described to each of the homogeneous groups separately. A variation on this possibility is to make the parameters of the utility function dependent on personal characteristics of individuals. The variation of behaviour across individuals is then exploited to estimate how the parameters of the utility function depend on these personal characteristics. In this thesis we adopt the latter procedure. We shall pay a fair deal of attention to a model which explains how the parameter δ varies across individuals according to family composition (below), and behaviour of other individuals, and own past circumstances (Chapter 4).

Somewhat arbitrarily, we model the influence of family composition by making the parameter δ dependent upon an indicator of the size of household k. This indicator may for instance comprise the number of family members, and the number of children under six. We have chosen for the following specification

$$\delta_k = \delta_0 + \delta_1 f_k \tag{2.5}$$

where f_k is the log of family size of household k
δ_0 and δ_1 are parameters.

The delta parameters are so-called translation parameters (cf. Pollak and Wales (1981)). To illustrate what this means, consider the utility function (2.3). It is immediately clear that this function can be written in general terms as

$$U(h_k, y_k) = F(h_k - \delta_k, y_k) \tag{2.6}$$

Now, consider for example a woman in two different situations (i.e with different parameters δ_1 and δ_2). Suppose she has more young children to take care of in situation 2 than in situation 1 and therefore: $\delta_1 > \delta_2$. In that case she will have a stronger taste for leisure in the second situation and this is illustrated by a lower delta. Let us assume that F is decreasing in h_k (more hours of work yields lower utility) and that F satisfies full comparability (Blackorby and Donaldson (1988)). As a result, for a given combination of y and h, she will be better off in situation 1 than in situation 2.

Both situations are illustrated in Figure 2.1. The two indifference curves represent exactly the same utility level for individual in the two situations. Relative to curve 1, curve 2 has been shifted to the right over a distance $\delta_1 - \delta_2$. This clearly indicates the stronger preference for leisure in situation 2. Given the (linear) budget constraint ABC, point B is optimal in situation 1. We see that she is worse off in situation 2 because indifference curve 2 (which would yield her the same utility level) is unattainable with this budget constraint, so she has to be satisfied with a lower level of utility.

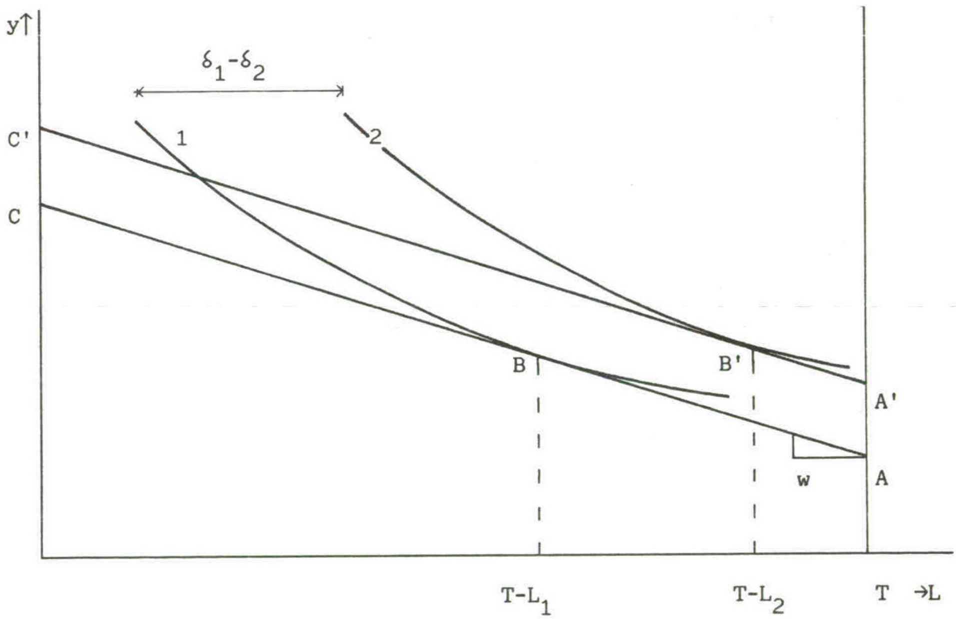


Figure 2.1 Two translated indifference curves

Incidentally, it is easy to construct a "money metric" for the utility difference between both individuals. The line A'B'C' has been drawn parallel to ABC. It is tangent to indifference curve 2 in point B'.

The points B and B' correspond to identical levels of consumption and to identical utility levels, but to different levels of leisure. The horizontal distance between B and B' equals $\delta_2 - \delta_1$. The budget constraints ABC and A'B'C' only differ with respect to the level of non-labour income. Apparently, she has to be compensated for her stronger preference for leisure by an additional amount of unearned income equal to AA' to achieve the same utility level as in situation 1.

Our choice to let family composition only affect the deltas implies that indifference curves can only shift sideways. This may be restrictive (although it still allows for shifts in both the intercept and the slope of the labour supply curves), but the tractability of the specification is an important reason to adhere to it.

An important issue is the identification of the parameters in equation (2.1) and (2.5). Clearly equation (2.1) is exactly identified. Substituting equation (2.5) into (2.1) and rewriting yields

$$h_k = \beta I_k + (\delta_0+\beta\vartheta) + (\gamma+\beta\delta_0)w_k + 1/2\gamma\beta w_k^2 + \delta_1 f_k + \beta\delta_1 f_k w_k \quad (2.7)$$

From equation (2.7) it is easy to see that this model is (over)identified.

Following common practice in estimating labour supply functions, a normally distributed error term with zero mean and variance σ_h^2 is added to the hours equation

$$h_k^o = h_k + \varepsilon_{hk} \quad \text{if} \quad h_k + \varepsilon_{hk} \geq 0$$
$$= 0 \qquad h_k + \varepsilon_{hk} < 0 \quad (2.8)$$

where $\varepsilon_{hk} \sim N(0, \sigma_h^2)$ (2.9)

h_k^o = observed number of hours individual k works.

The additive error term ε_{hk} could represent a measurement error in the hours variable or an optimization error on the part of the individual. Or it could represent deviations from the optimal number of hours due to demand side restrictions. In that case we have already deviated somewhat from a standard neoclassical labour supply model in the sense that it is no longer assumed that all unemployment is voluntary.

In fact, the whole model consists of 2 equations, the hours equation and the consumption equation. Because of adding-up restrictions the variance-covariance matrix of the error terms is singular. This problem can be accounted for by dropping arbitrarily one equation (cf. Barten (1969)). We have dropped the consumption equation.

2.3 The model specification for two-adult households

For ease of language we shall usually indicate the first two adults in a household as husband (or male) and wife (or female) respectively, and we will use two-adult households and families as synonyms. If there are any other adults in the household, their labour supply is taken to be exogenous just as the labour supply of children. The model to explain the joint labour supply decision of husband and wife is a generalization of the model for one-adult households, also referred to as single individuals. It is due to Hausman and Ruud (1984)

$$h_{mk} = \delta_{mk} + \gamma_m w_{mk} + \alpha w_{fk} + \beta_m I_k^* \tag{2.10}$$

$$h_{fk} = \delta_{fk} + \gamma_f w_{fk} + \alpha w_{mk} + \beta_f I_k^* \tag{2.11}$$

$$I_k^* = I_k + \vartheta + \delta_{mk} w_{mk} + \delta_{fk} w_{fk} + 1/2(\gamma_m w_{mk}^2 + \gamma_f w_{fk}^2) + \alpha w_{mk} w_{fk} \tag{2.12}$$

where h_{mk} = number of hours the male partner in household k would like to work per week

h_{fk} = number of hours the female partner in household k would like to work per week

w_{mk}, w_{fk} = after tax marginal wage rates of male and female, respectively

I_k = (weekly) nonlabour income of household k

ϑ, δ_{mk}, δ_{fk}, γ_m, γ_f, α, β_f, β_m are parameters.

The system (2.10)-(2.12) is quadratic in wages and hence exhibits a certain amount of flexibility in describing household labour supply. Clearly, a major assumption underlying this model is that the household is a homogeneous decision making unit, so that its preferences can be represented by a joint utility function. A different approach could be followed by placing the household decision making process in a bargaining framework. See, for example, theoretical papers by Manser and Brown (1980) or McElroy and Horney (1981) and empirical work by Bjorn and Vuong (1984) and Kooreman and Kapteyn (1985).

Under some regularity conditions the model is consistent with the maximization of a household utility function $U(h_{mk},h_{fk},y_k)$ under a linear budget constraint of the form

$$y_k = w_{mk}h_{mk} + w_{fk}h_{fk} + I_k \tag{2.13}$$

which states that the total amount of money available for consumption in a given period is the sum of unearned income I_k, labour income of the male $w_{mk}h_{mk}$ and labour income of the female $w_{fk}h_{fk}$. The functional form of $U(h_{mk},h_{fk},y_k)$ has been derived in Kapteyn, Kooreman and Van Soest (1989)

$$U(h_{mk},h_{fk},y_k) = \bar{I}_k \exp(\beta'\bar{w}_k) \tag{2.14}$$

where $\bar{I}_k = (\beta'A^{-1}\beta)^{-1} - [\, (\beta'A^{-1}\beta)^{-2} + (\beta'A^{-1}\beta)^{-1}$

$$\{(h_k - \delta_k)'A^{-1}(h_k - \delta_k) - 2(y_k + \vartheta)\}]^{1/2} \tag{2.15}$$

$$\bar{w}_k = A^{-1}(h_k - \delta_k - \bar{I}\,\beta) \tag{2.16}$$

$$\beta = \begin{bmatrix} \beta_m \\ \beta_f \end{bmatrix}, \quad A = \begin{bmatrix} \gamma_m & \alpha \\ \alpha & \gamma_f \end{bmatrix} \tag{2.17}$$

The concavity condition in this model is $(\beta'A^{-1}\beta)\bar{I}_k \leq 1$ for A positive definite (Kapteyn, Kooreman and Van Soest (1989)). The parameters δ_{mk} and δ_{fk} have been parameterized analogous to the model for single-adult households

$$\delta_{mk} = \delta_{m0} + \delta_{m1}f_k \tag{2.18}$$

$$\delta_{fk} = \delta_{f0} + \delta_{f1} f_k \tag{2.19}$$

The model described in equations (2.10)-(2.12) and (2.18)-(2.19) is clearly overidentified. To make this an estimable model, a stochastic specification is added to the model. This yields the following explanation of labour supply

$$\begin{aligned} h^o_{mk} &= h_{mk} + \varepsilon_{mk} \qquad \text{if } h_{mk} + \varepsilon_{mk} \geq 0 \\ &= 0 \qquad\qquad\qquad \text{if } h_{mk} + \varepsilon_{mk} < 0 \end{aligned} \tag{2.20}$$

$$\begin{aligned} h^o_{fk} &= h_{fk} + \varepsilon_{fk} \qquad \text{if } h_{fk} + \varepsilon_{fk} \geq 0 \\ &= 0 \qquad\qquad\qquad \text{if } h_{fk} + \varepsilon_{fk} < 0 \end{aligned} \tag{2.21}$$

$$\begin{bmatrix} \varepsilon_{mk} \\ \varepsilon_{fk} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_m & \rho\sigma_m\sigma_f \\ \rho\sigma_m\sigma_f & \sigma^2_f \end{bmatrix} \right) \tag{2.22}$$

where h^o_{mk} and h^o_{fk} are observed hours worked by husband and wife, h_{mk} and h_{fk} are given by equations (2.10)-(2.12) and ε_{mk} and ε_{fk} are error terms. Just as in the model for single-adult households the error terms can represent measurement errors, optimization errors, deviations from the optimal number of hours due to demand side restrictions.

2.4 Data

- OSA-Survey

The OSA-survey, which serves as one of the data-bases for estimation of the model, was held in 1985. Only persons between 18 and 60 years of age have been interviewed. In total the data set contains information on 4,020 individuals, of whom:

2,325 have a paid job;
177 are self-employed;
272 have no paid work, but are seeking;

1,243 have no paid work, and are not seeking;

3 are full time students.

For our model we need a fair amount of information per individual. As a result, a number of observations could not be used because the information for that individual was incomplete. Furthermore, we have excluded the self-employed from the sample, because the behaviour of this group appears to require a different model than the one set out in the previous sections. Disabled people, and students have been excluded as well, because they are not available for paid work in the short run. As a result the sample that is used for estimation consists of 849 households with two adults and 341 single-adult households. More detailed information on the composition of the sample is given in Table 2.1.

Sample means for some of the main variables of interest in this study, are given in Table 2.2. The wages reported are not only the actual wages observed in the sample, but also predicted wages. The reason for this is that in order to explain the behaviour of all households in the sample we need to know, also for those not working, what the potential wage rate is. Since no wage rate is observed for nonworking individuals a wage equation is estimated which explains an individual's wage rate as a function of his (or her) education and age. There are two ways of estimating this equation. The first method is to only use information about the wages of working individuals and estimate the equations by means of some method which corrects for the selectivity bias which may result from the fact that working and nonworking individuals may differ in certain unobservable but relevant aspects. The correction method used is Heckman's well known two step approach (Heckman(1979)). However, we have employed an alternative method, which exploits the fact that in the survey job seekers are asked the following question

> "If you could find the new job you are looking for, how many hours do you expect to work, and what do you think your net wage would be, taking into account your present chances in the labour market, and the kind of job you are looking for?"

The wage rate that nonworking seeking individuals report in response to the above question will be referred to as "expected wage rate". We assume that for nonworking individuals the expected after tax wage rate is the appropriate variable for the explanation of the labour supply decisions.

Thus we estimate for each level of education (log)wage equations for both workers and nonworkers together with the log of age and the squared log of age as explanatory variables. Estimation results are given in Appendix 2A. We don't present the estimation results obtained by the Heckman correction method, but it turned out that the sample means of the wage rates predicted according to the two different methods are not too far apart. It also turns out that for the estimation of the labour supply model it does not make a big difference which of the two wage equations is being used. See Kapteyn, Woittiez and ten Hacken (1989).

Regarding the sample means of actual hours, one observes that for males the number of hours worked per week seems to have gone up slightly from 1984 to 1985. This is not likely to be true, as during this period a strong political movement in the Netherlands has succesfully lobbied for a mandatory reduction of the working week. An explanation for the apparent rise in actual working hours may be that respondents tend to forget overtime hours they worked a year ago, so that the hours reported for last year are an underestimation. To the extent that this is a systematic effect it need not affect the estimation of the labour supply model too much, as it is mainly the variation across individuals in actual hours that serves to explain the current labour supply of households. The number of hours one prefers to work is also lower than the number of hours actually worked. For an extensive discussion of this variable one is referred to Chapter 3. Here it suffices to give the definition

> "Suppose you could freely choose the number of hours you work per week. How many hours would you like to work in your present job, if you could choose them yourself and if you would earn on average the same amount of money per hour as you do at the moment. If you choose fewer hours of work, you choose for less income. And more hours of work means more income. Assume that the number of hours of other members of the household, if any, do not change".

Next we turn to a graphical exposition of the distribution of wages and hours in the sample. Figure 2.2 gives the distribution of wages (including the distribution of predicted wages according to the method described above) of married males (married males and males in two-adult households are used as synonyms) and married females. First of all it is obvious, and as one might expect, that predicted wages vary less than actual wages.

This is a direct result of the fact that in the prediction the random part of the equation is ignored. Since wage equations are rather notorious for their poor statistical fit, ignoring the random part amounts to a sizeable reduction of observed variation, and this is exactly what the figures show. For the estimation of the linear labour supply model the rather poor fit of the wage equation has no dramatic implications, because the fit of the wage equation does not affect the consistency of the estimation of the parameters in the model. However, consistency of the estimated labour supply parameters is lost if wages not only appear linearly in the labour supply equation but also quadraticly. Despite this, it is common practice to impute wages into the labour supply equation. We will commit this error in Chapters 2 to 4, but in Chapter 5 and 6 we deal with the problem by estimating a simultaneous wage-hours model. Since the wage distributions of married males, single males, married females and single females all show the same features, we present only two of them. For the other wage distributions is referred to Kapteyn, Woittiez and ten Hacken (1988).

Comparing the wage distribution of males and females, it is clear that the distribution for the males is located more to the right than the distribution for females, in other words, generally males earn more than females. Comparing married males to single males shows that the distribution for the single males tends to be more to the left (single males earn less). For females the situation is rather the other way around. The higher wages of single females compared to those of married females may be explained by a difference in employment situation. The married females typically work part time and have experienced interruptions of their labour force participation (for example due to child bearing). These factors tend to reduce the wage one can earn.

Figure 2.3 presents distributions of actual and preferred hours for the same four groups as distinguished above. Considering the males first, we see that both the distributions for actual hours and for preferred hours show a spike at 40 hours a week. For single males there is a second spike at 32 hours. Generally the preferred hours distribution lies a bit more to the left than the actual hours distribution. The distributions for females are more dispersed with spikes at 20, 32, and 40 hours a week. For married females there is also a very large spike at 0 hours: more than 60% of the married females are not working. Also here the preferred hours distributions lie somewhat more to the left.

Table 2.1 Sample composition[a)]

male / female	working	nonworking nonseeking	nonworking seeking	all
working	315	3	13	331 (131)
nonworking, nonseeking	453	0	26	479 (36)
nonworking, seeking	33	0	6	39 (31)
all	801 (112)	3 (2)	45 (29)	849 (341=143+198)

a) The numbers in parentheses refer to singles

Table 2.2. Sample means

male	in families	single	all
actual hours per week	39.70	30.75	38.41
actual hours per week[a)]	42.07	39.26	41.73
actual hours per week, lagged 1 year	38.14	29.08	36.84
actual hours per week, lagged 1 year[a)]	39.50	34.63	38.90
preferred hours per week	36.49	28.08	35.28
preferred hours per week[a)]	38.68	35.86	38.33
actual net wage rate (guilders)[a)]	15.97	15.59	15.92
predicted net wage rate (guilders)[b)]	13.61	12.94	13.51
educ1 (1st level of education)	0.16	0.11	0.15
educ2 (2nd level of education)	0.20	0.15	0.19
educ3 (3rd level of education)	0.40	0.36	0.40
educ4 (4th level of education)	0.24	0.38	0.26
age	39.68	35.60	39.09
nonlabour income (guilders per week)	74.1	32.9	68.2
family size	3.46	1.10	3.12
seek[c)] (dummy=1 if individual is seeking, =0 if not)	0.94	0.94	0.94
unemployment benefit[d)]	333.6	261.5	303.5
number of all individuals	849	143	992
number of working individuals	801	112	913
number of nonworking individuals	48	31	79
number of individuals receiving unemployment benefit	42	30	72

a) Means based on working individuals.

b) Predicted wages, based on a wage regression for both working and nonworking individuals. Expected wages are used as observations for nonworking individuals. See Appendix 2A.

c) Means based on nonworking individuals.

d) Means based on nonworking individuals, receiving an unemployment benefit.

Table 2.2. Sample means, continued

female	in families	single	all
actual hours per week	10.64	23.63	13.10
actual hours per week[a)]	27.29	35.72	29.68
actual hours per week, lagged 1 year	11.96	22.05	13.87
actual hours per week, lagged 1 year[a)]	27.30	31.64	28.53
preferred hours per week	9.60	21.58	11.86
preferred hours per week[a)]	24.62	32.61	26.89
actual net wage rate (guilders)[a)]	12.54	14.13	12.99
predicted net wage rate (guilders)[b)]	11.13	11.31	11.17
educ1 (1st level of education)	0.26	0.18	0.24
educ2 (2nd level of education)	0.26	0.14	0.24
educ3 (3rd level of education)	0.38	0.44	0.39
educ4 (4th level of education)	0.10	0.24	0.13
age	37.16	36.20	36.98
nonlabour income (guilders per week)	74.1	121.0	83.0
family size	3.46	1.51	3.09
seek[c)] (dummy=1 if individual is seeking, =0 if not)	0.08	0.46	0.12
unemployment benefit[d)]	211.7	258.7	238.1
number of all individuals	849	198	1047
number of working individuals	331	131	462
number of nonworking individuals	518	67	585
number of individuals receiving unemployment benefit	12	48	60

a) Means based on working individuals.

b) Predicted wages, based on a wage regression for both working and nonworking individuals. Expected wages are used as observations for nonworking individuals. See Appendix 2A.

c) Means based on nonworking individuals.

d) Means based on nonworking individuals, receiving an unemployment benefit

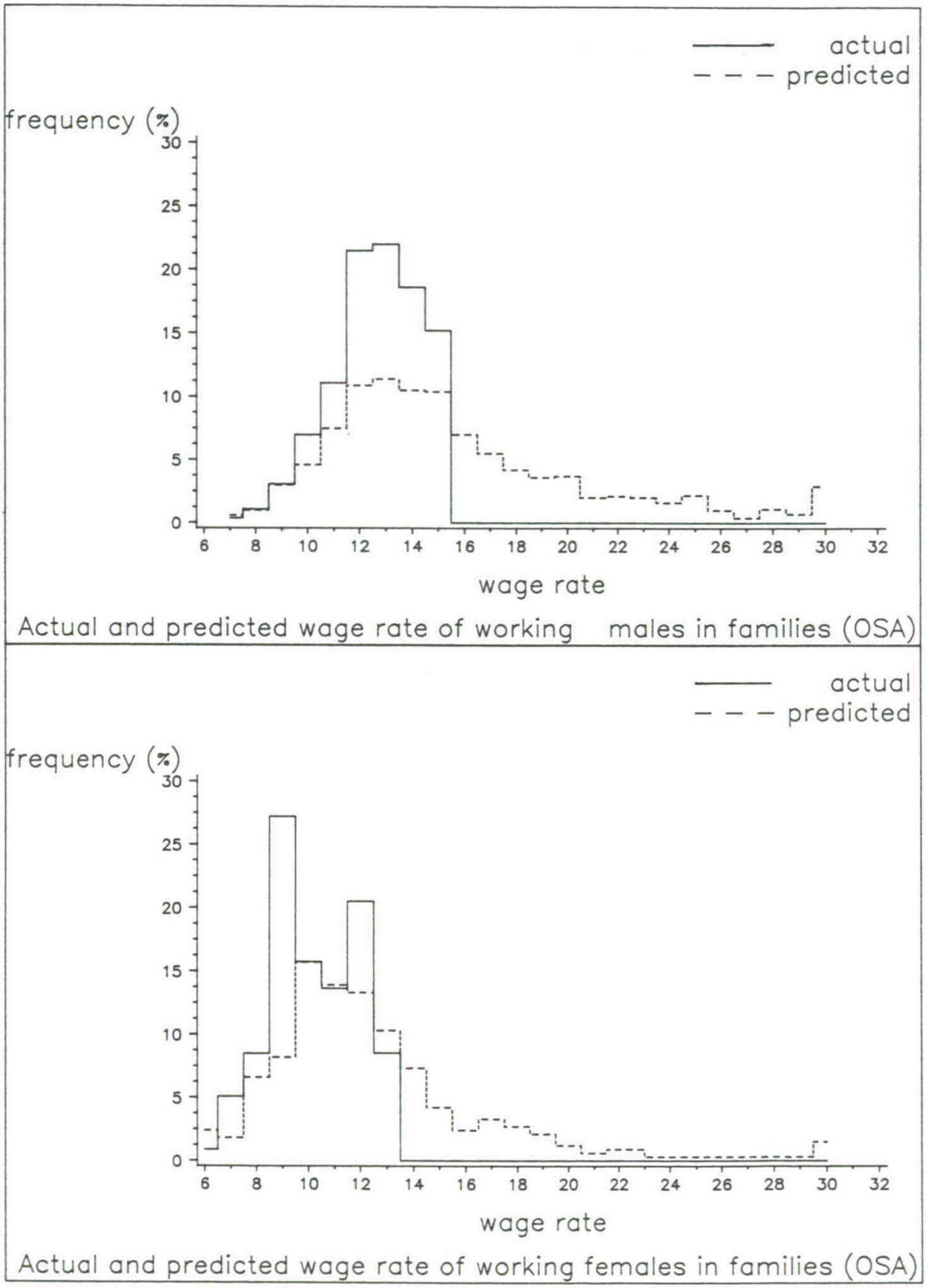


Figure 2.2 <u>Wage distributions for working males and females in families</u>

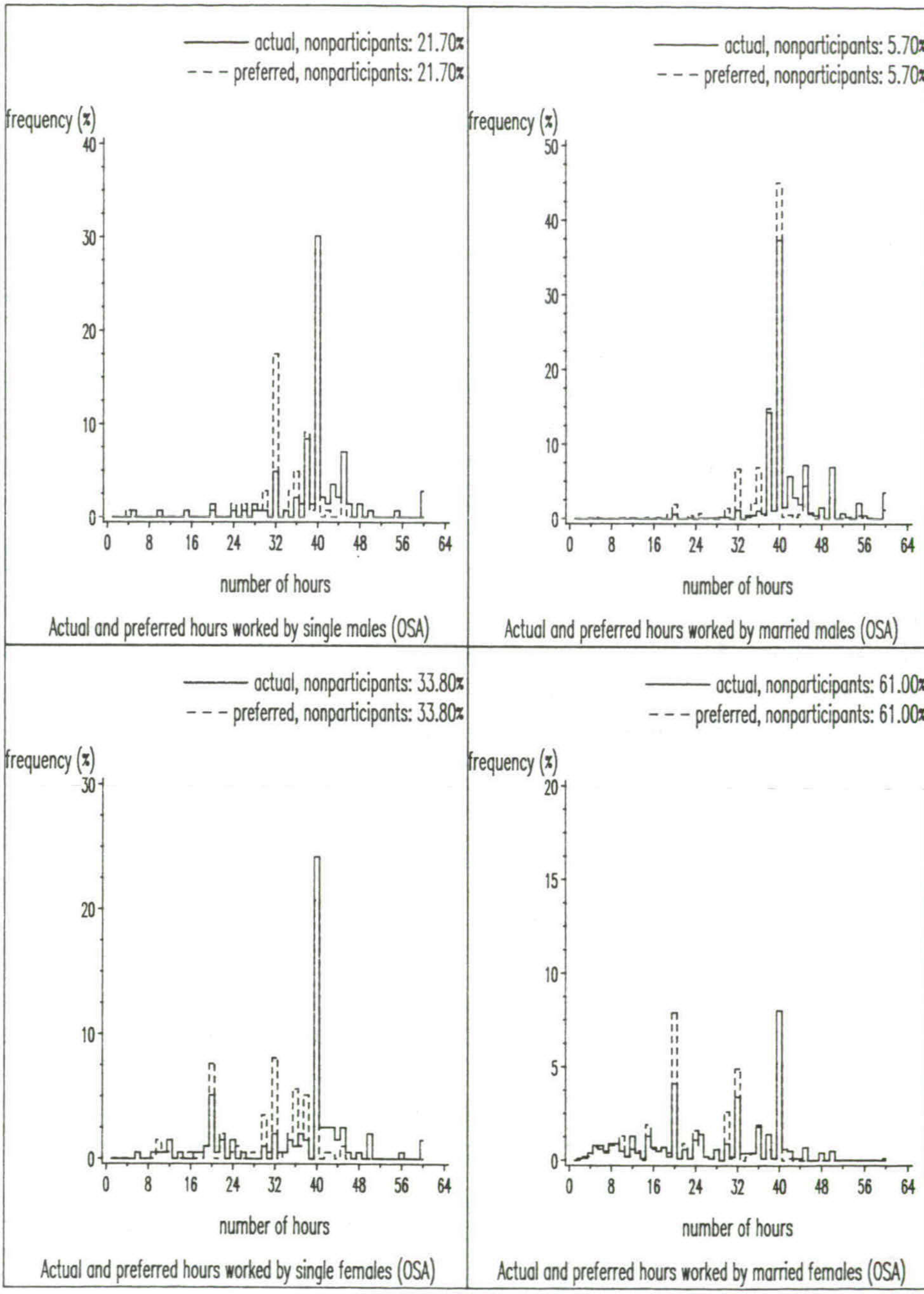


Figure 2.3 <u>Hours distributions</u>

- SEP-survey

The SEP is a bi-annual panel of households who are interviewed in April and October of each year. The panel started in 1984. For the estimation of the household labour supply models the wave of October 1985 has been used. The data set used contains 4225 households. In principle all persons in a household of 16 years or older have been interviewed. The information collected pertains primarily to incomes, labour market status, subjective evaluation of incomes, and various background variables like household composition, education, etc. It turns out that the income information for many households is incomplete, so that, even after a considerable amount of imputation, 358 records could not be used in estimation. Table 2.3 gives information on the characteristics of the sample that remains after the removal of the incomplete records and compares them with corresponding population information. Table 2.3 shows that compared with the population the sample contains fewer older people. In agreement with this, the percentage of widows and widowers in the sample is lower than in the population. Similarly, married couples are slightly overrepresented in the sample.

Taking this data set as a starting point of the analysis, we give its composition with respect to the employment status of the individuals in it. In total the data set contains 11881 individuals, of whom:

7237 (4020) are between 18 and 65 years old;
2805 (2325) have a paid job;
351 (177) are self-employed;
181 (272) have no paid work, but are seeking;
1908 (1243) have no paid work, and are not seeking;
30 (0) are in military service;
289 (3) are full time students.

The numbers in parentheses are the corresponding numbers for the OSA-survey. As with the OSA-survey, additional records had to be removed, since some other information needed for the estimation of the household labour supply model was lacking.

Table 2.3 The representativeness of the October '85-wave of the SEP

age	population[a] (%)	SEP(%)
0-19	28.3	31.3
20-44	39.5	39.7
45-64	20.3	19.3
65-79	9.4	8.5
80-	2.7	1.3

sex/ marital status	male		female	
	population[a] (%)	SEP(%)	population[a] (%)	SEP(%)
married	48.1	51.2	47.0	50.9
divorced	3.0	2.1	3.8	3.1
widowhood	2.1	1.7	9.0	6.4
unmarried	46.8	44.9	40.2	39.5
other	___	0.1	___	0.2
	49.5	49.8	50.5	49.8

type of household	population[a] (%)	SEP(%)
one-person household	28.6	20.0
no-family household	5.8	4.4
couple without children	20.9	23.7
couple with children	38.0	44.5
single parent family	5.2	5.3
other	2.5	2.0

a) See "Statistisch Zakboek", 1986.

The composition of the sample that was actually used for estimation, is given in Table 2.4. The table is analogous to Table 2.1.

Table 2.4 Sample composition[a)]

male / female	working	nonworking nonseeking	nonworking seeking	all
working	308	9	1	318 (85)
nonworking, nonseeking	589	44	10	643 (115)
nonworking, seeking	23	0	1	24 (12)
all	920 (117)	53 (32)	12 (22)	985 (383=171+212)

a) The numbers in parentheses refer to singles

It is rather striking that the SEP-sample used for estimation is approximately the same size as the OSA-sample, although the initial sample size of the SEP was about twice as large. The huge reduction of the sample size is mainly due to the large number of incomplete or unusable records that we have encountered in the data. The composition of both samples is roughly the same, except for one very striking fact, namely the percentage of nonseeking men. In the OSA-sample 45 out of the 48 nonworking males in families are seeking work and 29 out of 31 nonworking single males. In the SEP-sample the picture is quite different: 53 of the 65 nonworking males in families are nonseekers, and 32 of the 54 nonworking single males. All of these nonworking males receive some kind of benefit and the average age among this group is 52. This high age might explain why these individuals are not seeking.

Information on hours worked and wage rates in the SEP-sample is given in Table 2.5. Comparing Table 2.5 with Table 2.2 the differences between both samples appear to be small for males, but for females we see substantial differences. In the SEP-data the average number of hours

Information on hours worked and wage rates in the SEP-sample is given in Table 2.5. Comparing Table 2.5 with Table 2.2 the differences between both samples appear to be small for males, but for females we see substantial differences. In the SEP-data the average number of hours worked by married females is considerably less than in the OSA survey, this holds both for actual and preferred.

Like in the OSA-sample, the predicted wages vary less than the actual wages (see Figure 2.2). The wage distributions generally display the same features for both data sets:

- males earn more than females
- single males earn less than males in families
- single females earn more than females in families.

Figure 2.4 presents hours distributions. Comparing this figure with Figure 2.3 shows that once more spikes at 40 and 32 hours per week are present for males. For females there are also big spikes at 20 and 0 hours.

Table 2.5. Sample means

male	in families	single	all
actual hours per week	39.41	26.47	37.49
actual hours per week[a)]	42.19	38.69	41.80
actual hours per week, lagged 1 year	39.09	25.84	37.13
actual hours per week, lagged 1 year[a)]	41.43	36.93	40.92
preferred hours per week	38.23	32.85	37.44
preferred hours per week[a)]	39.52	35.37	39.05
actual net wage rate (guilders)[a)]	14.50	13.22	14.35
predicted net wage rate (guilders)[b)]	13.60	12.48	13.44
educ1 (1st level of education)	0.14	0.20	0.15
educ2 (2nd level of education)	0.18	0.25	0.19
educ3 (3rd level of education)	0.47	0.36	0.46
educ4 (4th level of education)	0.21	0.18	0.21
age	39.93	37.90	39.63
nonlabour income (guilders per week)	102.2	165.0	111.5
family size	3.49	1.50	3.20
seek[c)] (dummy=1 if individual is seeking, =0 if not)	0.18	0.41	0.29
unemployment benefit[d)]	418.9	290.3	350.7
number of all individuals	985	171	1156
number of working individuals	920	117	1037
number of nonworking individuals	65	54	119
number of individuals receiving unemployment benefit	39	44	83

a) Means based on working individuals.

b) Predicted wages, based on a wage regression for both working and nonworking individuals. Expected wages are used as observations for nonworking individuals. See Appendix 2A.

c) Means based on nonworking individuals.

d) Means based on nonworking individuals receiving an unemployment benefit

Table 2.5. Sample means, continued

female	in families	single	all
actual hours per week	7.31	13.58	8.42
actual hours per week[a)]	22.65	33.88	25.02
actual hours per week, lagged 1 year	7.76	13.80	8.83
actual hours per week, lagged 1 year[a)]	21.75	32.91	24.10
preferred hours per week	10.00	17.04	11.25
preferred hours per week[a)]	21.70	31.66	23.80
actual net wage rate (guilders)[a)]	12.54	12.87	12.61
predicted net wage rate (guilders)[b)]	11.17	11.19	11.17
educ1 (1st level of education)	0.19	0.32	0.22
educ2 (2nd level of education)	0.34	0.18	0.31
educ3 (3rd level of education)	0.36	0.32	0.35
educ4 (4th level of education)	0.10	0.18	0.12
age	37.34	44.87	38.67
nonlabour income (guilders per week)	102.2	216.4	122.4
family size	3.49	1.62	3.16
seek[c)] (dummy=1 if individual is seeking, =0 if not)	0.04	0.09	0.05
unemployment benefit[d)]	211.7	262.4	255.7
number of all individuals	985	212	1197
number of working individuals	318	85	403
number of nonworking individuals	667	127	794
number of individuals receiving unemployment benefit	12	78	90

a) Means based on working individuals.

b) Predicted wages, based on a wage regression for both working and nonworking individuals. Expected wages are used as observations for nonworking individuals. See Appendix 2A.

c) Means based on nonworking individuals.

d) Means based on nonworking individuals, receiving an unemployment benefit

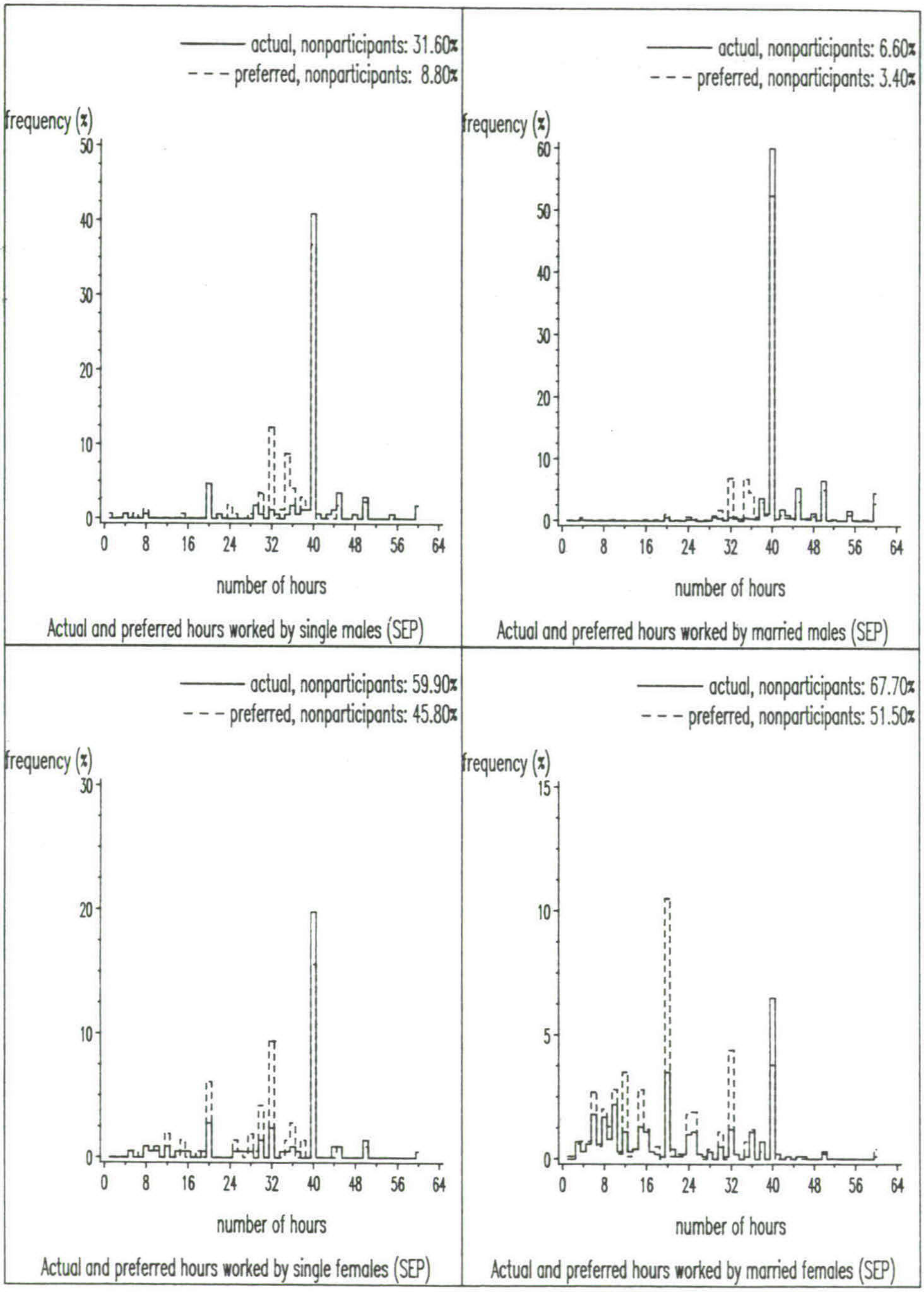


Figure 2.4 Hours distributions

2.5 Estimation Results

- One-adult households

The exact specification of the model is given in Section 2.2, equations (2.7) and (2.8). As endogenous variable we use the actual number of hours worked. By using the average wage rate instead of the marginal wage rate it is implicitly assumed that the individual faces a linear budget constraint. Of course this is a simplification of reality, but it turns out that in practice this assumption doesn't alter the results very much. In Chapter 3 we will nevertheless relax this assumption. In estimation predicted wages are used both for working and not working individuals. The wage equations used for prediction are given in Appendix 2A. We have excluded individuals receiving unemployment benefits from the sample.

The appropriate likelihood contribution of working individuals is:

$$L^1 = \frac{1}{\sigma_h} \varphi\left[\frac{h^o - h}{\sigma_h} \right] \tag{2.23}$$

where φ is the standard normal density function.

For nonworking individuals in the data sets used, it is known whether they are looking for a job or not. If one is looking for a job this means in terms of (2.1) that h_k is positive, and if one is not seeking, h_k is less than or equal to zero. The likelihood contribution of an individual who is not working and not looking for a job is:

$$L^2 = \Phi\left[\frac{-h}{\sigma_h} \right] \tag{2.24}$$

where Φ is the standard normal distribution function.
The contribution of a non-working job seeker is:

$$L^3 = \Phi\left[\frac{h}{\sigma_h} \right] \tag{2.25}$$

The estimation results are given in Table 2.6. Let us first concentrate on the results based on the OSA-data. Looking at the labour supply equation for males, it emerges that neither nonlabour income nor family size nor the wage rate are significant. According to these estimation results, male labour supply is explained by a constant of about 37 hours per week. These results are confirmed by a reduced form estimation where we have regressed male hours per week on a constant, (the deviation in) the wage rate, (the deviation in) the squared wage rate and the (deviation in) family size. The estimation results of this regression are presented in Appendix 2B. For females almost all coefficients are significant, except for family size and the constant term. The lower panel of Table 2.6 is based on the SEP-data. For males the income coefficient is significantly negative implying leisure to be a normal good, all other coefficients are insignificant, while for females the wage rate has a positive effect and nonlabour income a negative effect on hours worked, both significant. The coefficient ϑ is fixed since we ran into numerical identification problems, due to the small estimated value of β (note that ϑ cannot be identified if β=0). One should realize that all other parameter estimates and their estimated standard errors are conditional on the fixed value of ϑ. The parameter estimates based on the OSA-sample differ quite a bit from the estimates based on the SEP-sample. This is not surprising in the light of the differences in the sample composition. For example, 13.1% of the single females in the OSA-sample, and 38.9% of the single females in the SEP-sample do not participate (Compare Table 2.2 and 2.5). Note that we use nonparticipants and nonworkers as synonyms ,contrary to what is common in the literature. In Chapter 7 we will test two data sets against each other.

Table 2.6 Estimation results for the standard model[a)]

OSA

	males		females	
δ_0	37	(6)	4	(10)
β	-0.000001	(u.b.)	-0.041	(0.005)
δ_1	0.6	(6)	-8	(5)
γ	0.1	(0.5)	2.6	(0.6)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	11.7	(0.5)	15.5	(1.1)
log likelihood				
	-430.9		-542.4	

SEP

	males		females	
δ_0	14	(10)	-24.6	(8.2)
β	-0.0091	(0.003)	-0.067	(0.008)
δ_1	4.8	(3.8)	4	(14)
γ	0.8	(0.6)	2.6	(0.4)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	16.0	(1.0)	22.0	(1.8)
log likelihood				
	-486.5		-380.8	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses

- Two-adult households

Equations (2.10)-(2.12) and (2.20)-(2.22) specify the model that has been estimated for two-adult households. The likelihood function consists of different parts (L_k) corresponding to the different situations households face. Let Φ and φ be the standard normal distribution and density function respectively. Let $B\Phi$ and $b\varphi$ be the bivariate standard normal distribution and density functions, respectively (with correlation ρ).

We distinguish the following situations (where i,j stands for m(ale) or f(emale)):

1) both spouses are working ($h^o_{ik} > 0$, $h^o_{jk} > 0$)

$$L^1_k = \frac{1}{\sigma_i \; \sigma_j} \, b\varphi\left[\frac{h^o_{ik} - h_{ik}}{\sigma_i}, \frac{h^o_{jk} - h_{jk}}{\sigma_j}, \rho\right] \tag{2.26}$$

2) spouse i is not working and not seeking a job, and spouse j is working ($h^o_{ik} = 0$, $seek_{ik} = 0$, $h^o_{jk} > 0$)

$$L^2_k = \Phi\left[\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left[h^o_{jk} - h_{jk}\right]}{\sigma_j\sqrt{1-\rho^2}}\right] \frac{1}{\sigma_j} \, \varphi\left[\frac{h^o_{jk} - h_{jk}}{\sigma_j}\right] \tag{2.27}$$

3) spouse i is not working but seeking a job, and spouse j is working ($h^o_{ik} = 0$, $seek_{ik}=1$, $h^o_{jk} > 0$)

$$L^3_k = \left[1-\Phi\left[\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left[h^o_{jk} - h_{jk}\right]}{\sigma_j\sqrt{1-\rho^2}}\right] \right] \frac{1}{\sigma_j} \, \varphi\left[\frac{h^o_{jk} - h_{jk}}{\sigma_j}\right] \tag{2.28}$$

4) Both spouses are not working and not seeking a job ($h^o_{ik} = 0$, $seek_{ik}=0$, $h^o_{jk} = 0$, $seek_{jk}=0$)

$$L_k^4 = B\Phi\left[\frac{-h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, \rho\right] \tag{2.29}$$

5) Both spouses are not working and i is not seeking a job, but j is seeking ($h_{ik}^o = 0$, $seek_{ik}=0$, $h_{jk}^o = 0$, $seek_{jk}=1$)

$$L_k^5 = B\Phi\left[\frac{-h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, -\rho\right] \tag{2.30}$$

6) Both spouses are not working but seeking a job ($h_{ik}^o = 0$, $seek_{ik}=1$, $h_{jk}^o = 0$, $seek_{jk}=1$)

$$L_k^6 = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right] \tag{2.31}$$

where $seek_{ik}$ = 0 if spouse i in household k is not seeking a job
= 1 otherwise.

The first column of Table 2.7 corresponds with the OSA-data set, the second with the SEP. We note that both columns of Table 2.7 show negative income effects (β_m and β_f), implying leisure to be a normal good, and positive own linear wage effects (γ_m and γ_f). Note however that for the SEP-sample the male wage coefficient and the female income coefficient had to be set at the lower and upper bound, respectively and for the OSA-sample most wage and income effects are insignificant. Furthermore, we see that in both models family size has a negative effect on the female's labour supply. In fact the first child reduces the number of working hours of the wife by 19 (17), the second child by 14 (12) more in the OSA-sample (SEP).

Table 2.7 Estimation results for the standard model[a)]

	OSA		SEP	
α	-0.1	(2.8)	-0.02	(0.15)
β_m	-0.002	(0.002)	-0.021	(0.001)
β_f	-0.022	(0.006)	-0.000001	(u.b.)
γ_m	0.1	(0.3)	0.01	(ℓ.b.)
γ_f	2	(26)	4.1	(0.6)
δ_{m1}	-0.2	(1.4)	-6.5	(1.9)
δ_{f1}	-61.5	(8.1)	-41.8	(3.6)
δ_{m0}	33	(89)	-11.9	(5.1)
δ_{f0}	-31	(828)	-4	(6)
ϑ	-3183	(38977)	-2436	(192)
σ_m	6.88	(0.08)	9.4	(0.1)
σ_f	24.0	(1.6)	24.5	(1.6)
ρ	-0.03	(0.04)	0.05	(0.04)
log likelihood	-4384.2		-4971.0	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses

In Table 2.8 own wage elasticities for both one-adult and two-adult households are given.

Table 2.8 Wage elasticities for the standard model[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m} \frac{w_m}{h_m}$	0.04	-0.00	0.20	0.14
	(0.16)	(0.06)	(0.21)	(0.04)
$\frac{\partial h_f}{\partial w_f} \frac{w_f}{h_f}$	0.43	2.04	0.91	4.72
	(0.24)	(3.92)	(0.31)	(4.22)

a) Standard errors in parentheses

The elasticities in Table 2.8 are aggregate elasticities. They measure the responsiveness of the total labour supply of all individuals to a one percent increase in each individual's wage rate. This is different from calculating the elasticity in the sample mean. Unfortunately, it is not common practice to present standard errors of elasticities. But especially if the standard errors of the parameters are large, one expects the standard errors of the elasticities also to be large. In Appendix 2C the derivation of the standard errors presented in parentheses in Table 2.8 is given. A couple of remarks can be made with respect to this table. The wage elasticities for males are all low and not significantly different from 0, except for married males in the SEP. The wage elasticities for single females are somewhat larger and significantly different from 0. For married females the elasticities are much larger but their standard errors are large too. It is striking to note the large differences between the OSA and SEP estimates of the elasticities. In all cases the SEP estimates are higher than the OSA estimates. Whether this tendency is found for other specifications as well will become clear in Chapter 7 of this thesis.

In Figures 2.5 and 2.6 the labour supply curves are shown. The figures are obtained as follows: Each individual's wage rate is varied by 10% steps from -25% to +25% of the original wage rate. For each of the different wage rates and for each individual the deterministic part of the optimal number of hours is simulated. These numbers are then averaged over the sample of working individuals. One should note that the labour supply wage elasticities and the labour supply curves display different information. The elasticities are computed as aggregates, based on all individuals. The labour supply curves are based on means of working individuals only. As one can see mean male labour supply varies from 38 to 40 hours per week for singles and is constant at about 42 hours per week for males in families (OSA-sample). In the SEP-sample male labour supply varies a bit more between 35 and 40 hours for single males and from 39 to 43 for males in families. Both male labour supply curves in two-adult households are higher than the corresponding curves for single males. As one would expect, the female labour supply curves of singles lie well above the supply curve of married females. The latter lies between 12 and 18 hours per week in the OSA sample and between 5 and 15 hours in the SEP. The OSA female labour supply curves lie higher than the SEP curves. The figures show quite clearly that female labour supply is a quadratic function of the wage rate, while for males the labour supply function is almost linear in wages.

Finally, Figures 2.7 and 2.8 display the actual and simulated hours distributions. The simulated hours distributions are spread out by the stochastic term, and are not able to explain the spikes in the actual hours distributions. The simulated hours distributions of males show small thick peaks around 40, while the actual hours distributions show large thin peaks at 40. The model cannot reproduce the 8% nonparticipants of single males in the SEP; the simulated percentage of nonworkers is 1.2. The peaks in simulated hours at 64 hours per week arise because we have cut off the distributions at 64. The peaks in the actual female hours distributions at 0 are all reproduced, although in all cases the degree of nonparticipation is underestimated. But the other small peaks at 20 and 40 hours cannot be reproduced. Due to the high actual degree of nonparticipation of married females the peak in the simulated hours distribution is at 0.

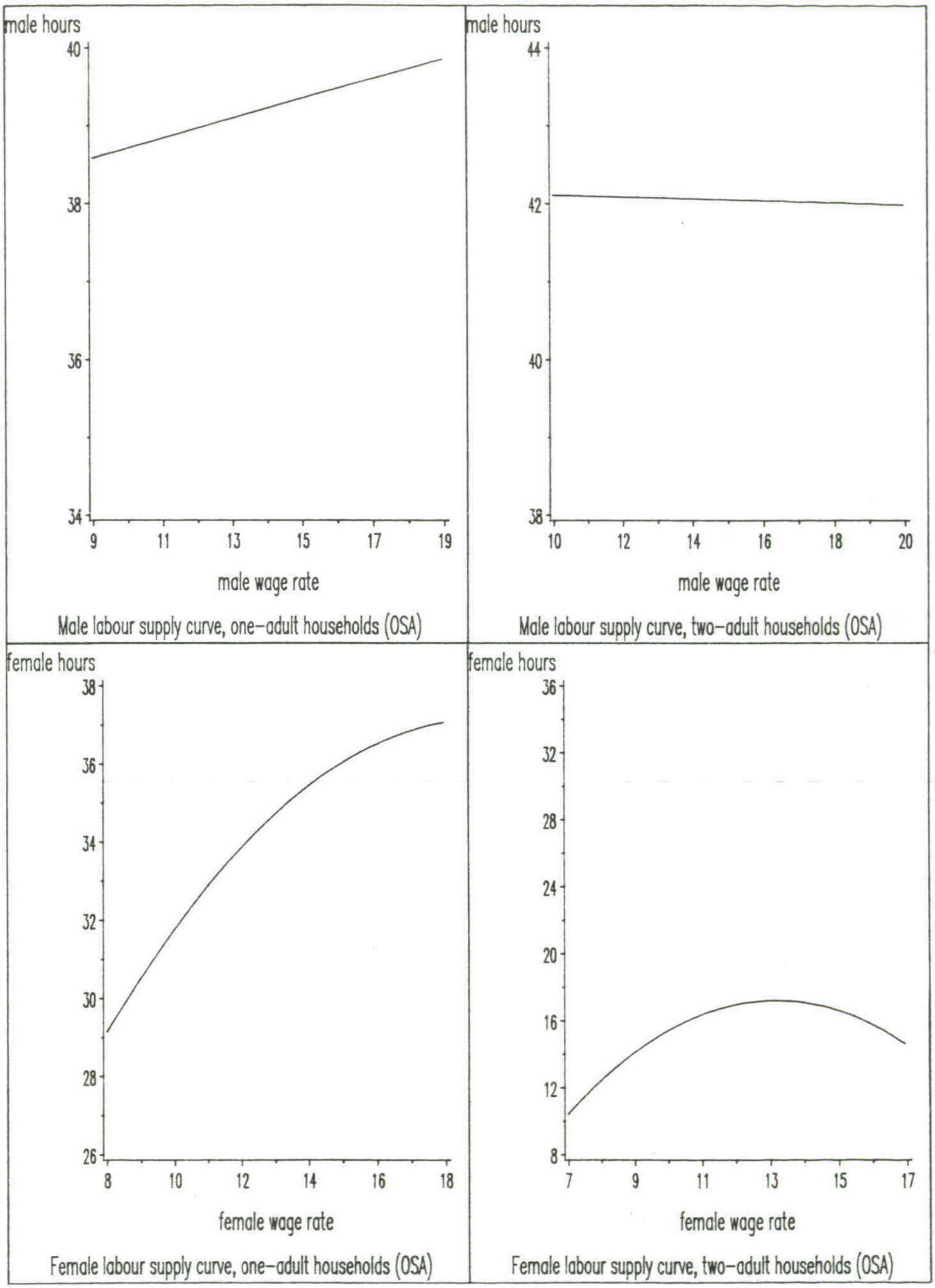


Figure 2.5 Labour supply curves (OSA)

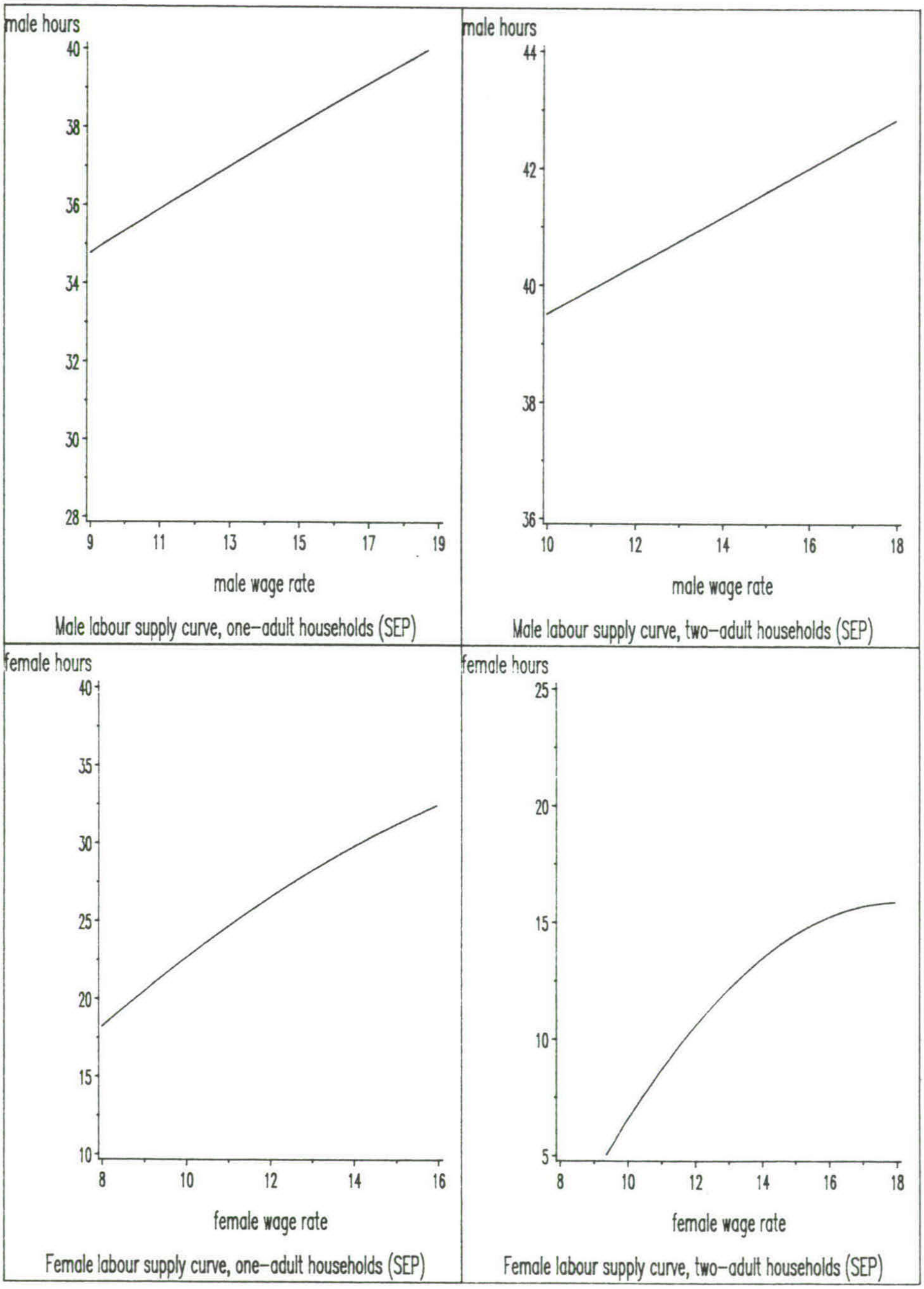


Figure 2.6 Labour supply curves (SEP)

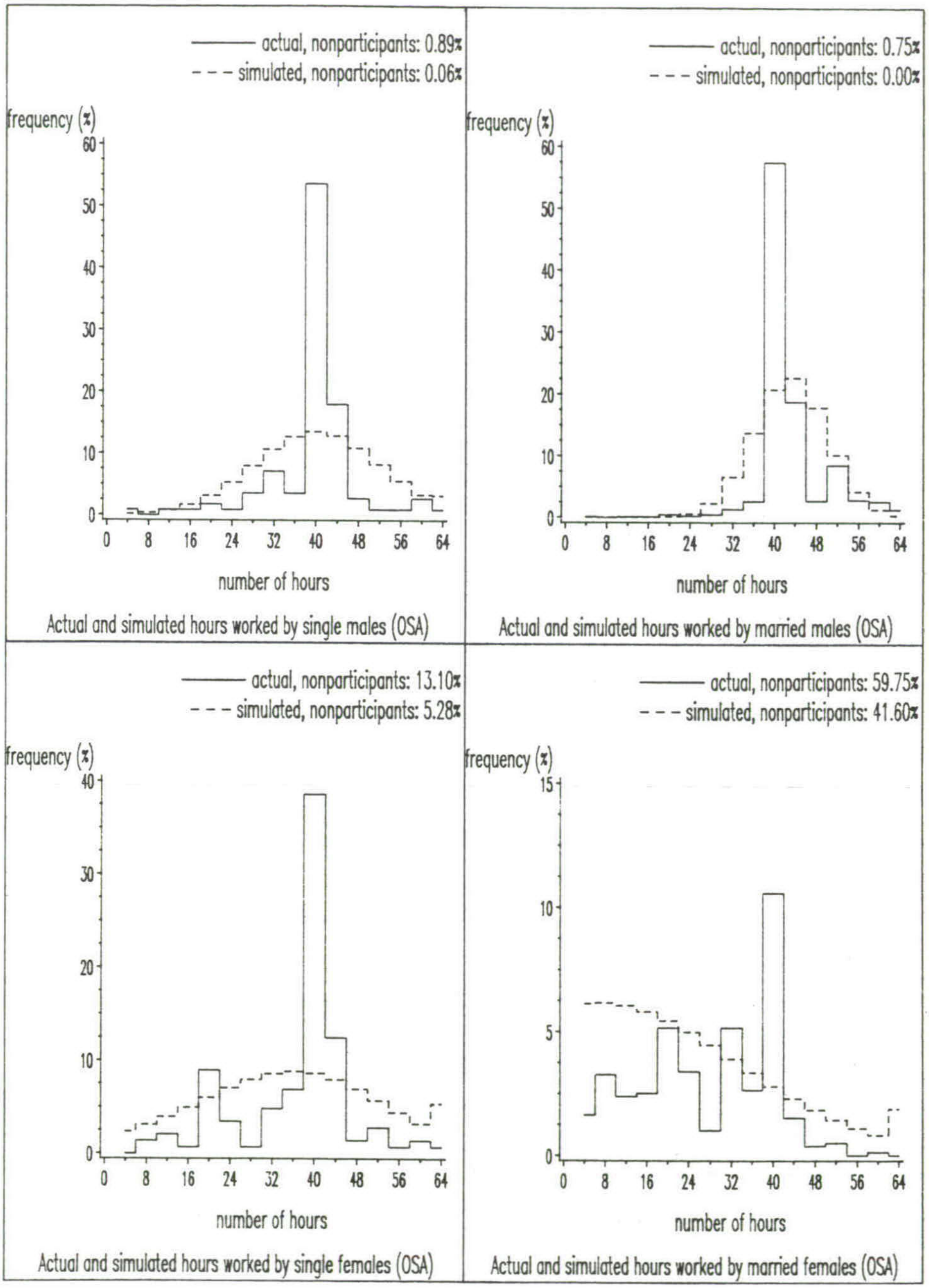


Figure 2.7 Hours distributions (OSA)

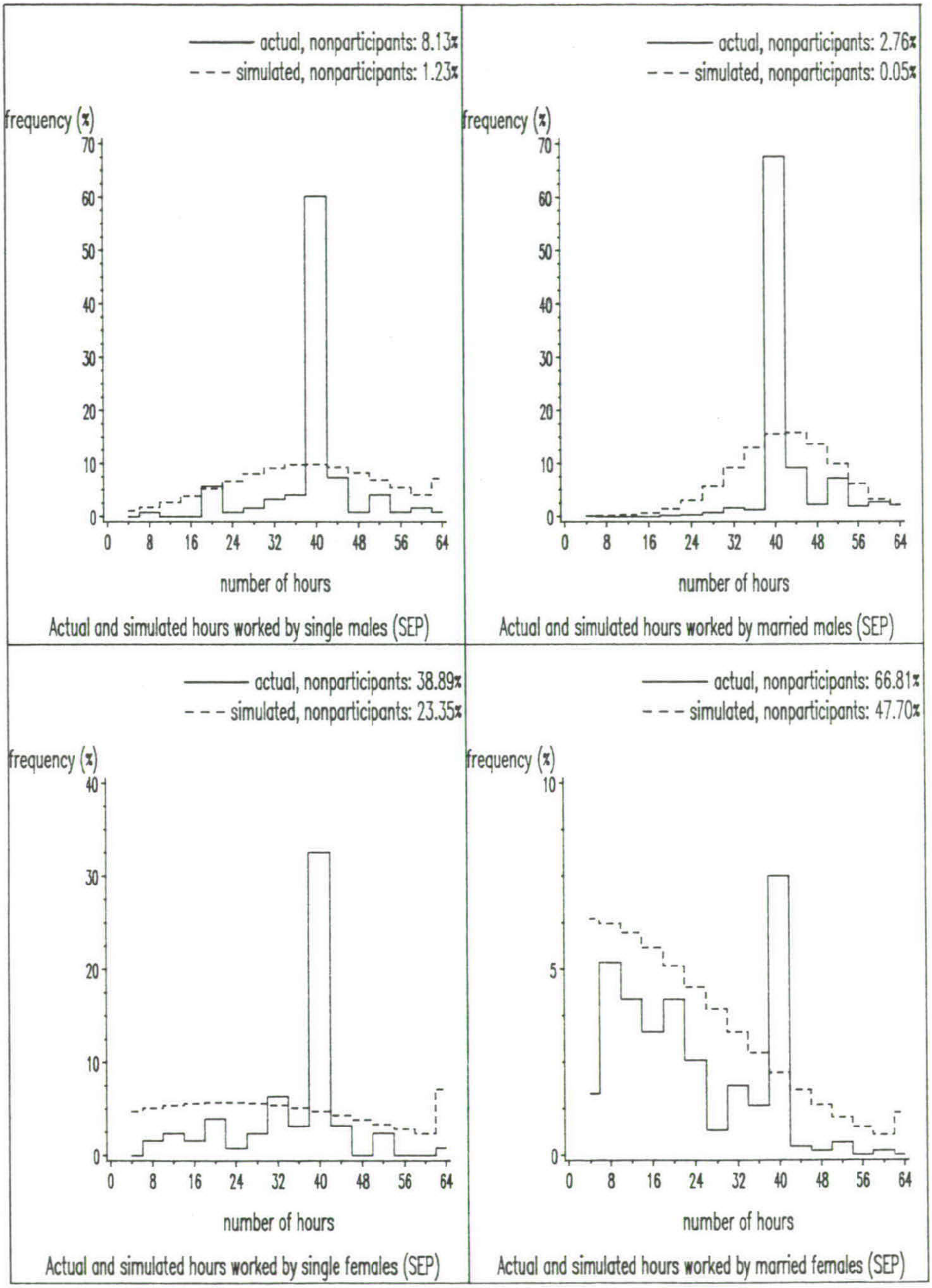


Figure 2.8 Hours distributions (SEP)

2.6 Concluding remarks

This chapter has two major points of emphasis. The first is the modelling of a standard neoclassical labour supply model for one-adult households (Section 2.2) and two-adult households (Section 2.3). The second is the extensive description of the two data sets: Namely the OSA-data set and the SEP-data set (Section 2.4). In Section 2.5 some estimation results are shown for simple neoclassical models of labour supply. The use of different data sets yields results that are rather different as far as wage elasticities, hours distributions and labour supply curves are concerned, although in each case we find low own wage elasticities for males and higher elasticities for females. Moreover, we find that the presence of children has a strong negative effect on the labour supply of females in families.

Appendix 2A Wage equations

The data sets we use do not only contain actual net wage rates for working individuals, but also expected net wage rates for nonworking individuals. We have estimated a wage equation using both types of information. Estimation results for the OSA-data are presented in Table 2A.1, the results for the SEP can be found in Table 2A.2.

From Table 2A.1 it can be concluded that the higher the education the higher the age for which the male wage-age curve takes its maximum: The age at which the maximum is attained increases from 41 for the lowest level of education, via 43 and 52 to 61 for the highest level of education. In contrast females generally receive the highest wage rate when they are about 40 years old, irrespective of level of education. In Table 2A.2 (SEP) the same phenomena are present, except that both men and women with the highest level of education do not reach the top of the wage-age profile (the corresponding age is well above 65).

Table 2A.1: Wage-equations, based on actual and expected wage-rates[a),b)] (OSA-data)

Log wage-equation for men

level of education	constant	log(age)	$[\log(age)]^2$	number of observations	R^2
1	-9.34	6.40	-0.86	290	0.11
	(2.13)	(1.29)	(0.18)		
2	-12.60	8.12	-1.08	340	0.27
	(2.06)	(1.17)	(0.17)		
3	-7.72	5.30	-0.67	656	0.16
	(1.94)	(1.08)	(0.15)		
4	-6.01	4.27	-0.52	385	0.16
	(2.75)	(1.53)	(0.21)		

log wage-equation for women

level of education	constant	log(age)	$[\log(age)]^2$	number of observations	R^2
1	-6.11	4.46	-0.59	250	0.12
	(2.30)	(1.31)	(0.19)		
2	-10.46	7.04	-0.96	280	0.18
	(2.28)	(1.31)	(0.19)		
3	-8.41	5.85	-0.78	540	0.15
	(1.79)	(1.03)	(0.15)		
4	-7.98	5.65	-0.75	194	0.09
	(4.11)	(2.31)	(0.32)		

a) Standard errors in parentheses.

b) Education has been coded in 4 levels, ranging form 1 (lowest) till 4 (highest).

Table 2A.2 Wage-equations, based on actual and expected wage-rates[a),b)] (SEP-data)

Log wage-equation for men

level of education	constant	log(age)	[log(age)]^2	number of observations	R^2
1	-2.58	2.68	-0.36	141	0.02
	(4.01)	(2.21)	(0.30)		
2	-2.01	2.35	-0.31	214	0.03
	(2.83)	(1.58)	(0.22)		
3	-8.06	5.66	-0.75	589	0.07
	(2.22)	(1.22)	(0.17)		
4	-3.13	2.79	-0.31	270	0.11
	(5.01)	(2.71)	(0.37)		

log wage-equation for women

level of education	constant	log(age)	[log(age)]^2	number of observations	R^2
1	-14.1	9.27	-1.29	68	0.07
	(7.41)	(3.40)	(0.56)		
2	-19.8	12.6	-1.77	138	0.07
	(6.8)	(3.8)	(0.54)		
3	-9.07	6.34	-0.87	192	0.07
	(4.03)	(2.30)	(0.33)		
4	-0.87	1.54	-0.15	78	0.05
	(11.20)	(6.29)	(0.88)		

a) Standard errors in parentheses.

b) Education has been coded in 4 levels, ranging form 1 (lowest) till 4 (highest).

Appendix 2B A reduced form estimation of single male labour supply

The results presented in Table 2.6 seem to indicate that male labour supply is best explained by a constant of about 37 hours per week. It could be the case that the functional form we have imposed is too restrictive. Therefore, we have estimated the following reduced form equation instead of equation (2.1) by means of maximum likelihood

$$h_k = c_1 + c_2(I_k-\bar{I}) + c_3(w_k-\bar{w}) + c_4(w_k^2-\overline{w^2}) + c_5(f_k-\bar{f}) + c_6(w_kf_k-\overline{wf}) \quad (2B.1)$$

where a bar (-) on a variable denotes its sample mean.

We have substracted the mean of each variable to minimize possible numerical problems. In Table 2B.1 the estimation results are presented. The near equality of the two likelihood values in Table 2.6 and Table 2B.1 suggests that the imposed functional form is not too restrictive.

Table 2B.1 Estimation results for the standard model[a)]

	OSA	
	males	
c_1	41	(7)
c_2	-0.000001	(u.b.)
c_3	11	(6)
c_4	-0.4	(0.2)
c_5	-0.3	(116)
c_6	0.3	(8.7)
σ_h	11.6	(0.5)
log likelihood		
	-429.8	

Appendix 2C Derivation of the standard errors of elasticities

In this Appendix is shown how the standard errors of an elasticity can be derived. The elasticities shown in Table 2.8 are aggregated over the individuals in the sample. The aggregate wage elasticity of single males, for example, can be written as

$$\text{elas} = \frac{\sum_{k=1}^{K} [h_k(w_k+\Delta w_k)-h_k(w_k)]}{\sum_{k=1}^{K} \Delta w_k} \quad \frac{\sum_{k=1}^{K} w_k}{\sum_{k=1}^{K} h_k} \tag{2C.1}$$

where h_k from equation (2.1) is written as a function of w_k: $h_k(w_k)$ and $\Delta w_k = 1.01 w_k$. This elasticity measures the total increase (or decrease) in hours worked of all individuals in the population as a result of a 1% increase in each person's wage rate.
From (2.1) we know that h_k is a nonlinear function of the parameter vector Θ and exogenous variables X_k:

$$h_k = g_1(\Theta, X_k) \tag{2C.2}$$

where $\Theta = (\delta_0 \;\; \delta_1 \;\; \gamma \;\; \beta \;\; \vartheta)$ and

$$X_k = (w_k \;\; I_k \;\; f_k)$$

Likewise the aggregate elasticity can be written as a function of Θ

$$\text{elas} = g_2(\Theta, X) \tag{2C.3}$$

The parametervector Θ has been estimated by means of maximum likelihood. Therefore

$$\sqrt{K}\ (\hat{\Theta}-\Theta) \overset{A}{\sim} N\ (0,\ \Sigma) \tag{2C.4}$$

Then the asymptotic distribution of elas is

$$\sqrt{K}\ (g_2(\hat{\Theta},X)-g_2(\Theta,X)) \overset{A}{\sim} N(0,\ Vg_2) \tag{2C.5}$$

$$Vg_2 = \sum_{i=1}^{I} \sum_{j=1}^{I} \left[\frac{\partial}{\partial \vartheta_i} g_2(\Theta,X) \Big|_{\Theta} \right] \left[\frac{\partial}{\partial \vartheta_j} g_2(\Theta,X) \Big|_{\Theta} \right] \Sigma_{ij} \qquad (2C.6)$$

where I is the number of parameters

ϑ_i is a typical element of the parameter vector Θ

Σ_{ij} is a typical element of Σ.

3 The social security and welfare system, and institutional constraints

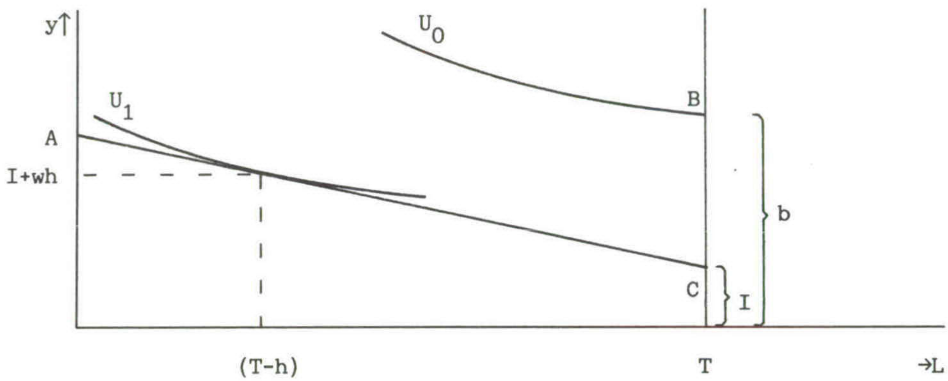
3.1 Introduction

Generally, the budget constraint will not have the linear form that was assumed in Chapter 2. Due to the tax and the social security and welfare system, it will be nonlinear and nonconvex. In Section 3.2 we will concentrate on modelling labour supply if the budget set is nonconvex, the nonconvexity being caused by the Dutch social security and welfare system.

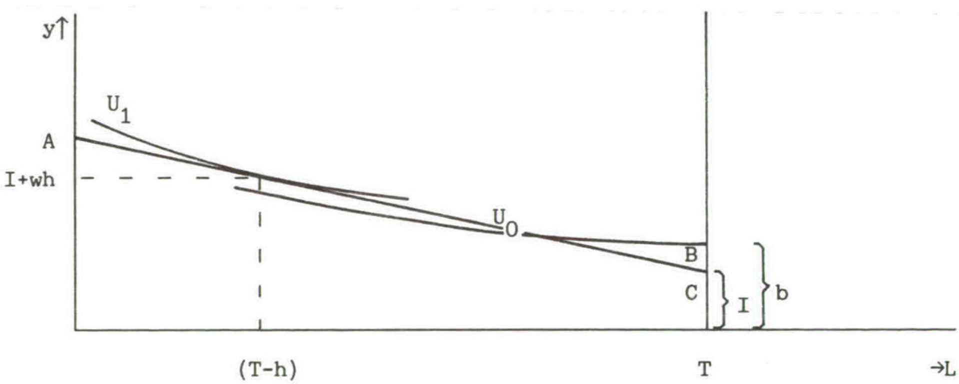
It is common practice to use actual hours worked as the endogenous variable in labour supply models. But actual hours are generally the result of the interplay of household preferences, institutional constraints and the demand for labour. In Chapter 6 of this thesis we will deal explicitly with demand side constraints. In Section 3.3, however, a model of labour supply is presented in which the problem of dealing with demand side constraints has been circumvented. For that purpose we use a specific question in both the OSA and SEP survey about how many hours an individual would like to work. Rationing theory has to be used due to the wording of the question (see Kapteyn, Kooreman and van Soest (1989)). In Section 3.4 estimation results are presented and Section 3.5 concludes.

3.2 A nonconvex budget set

For the moment we abstract from the nonlinearities cuased by the tax system. Let us concentrate on the Dutch social security system. If a presently unemployed person receives unemployment benefits, and this person takes a job, his or her benefits wil be reduced. This introduces nonconvexities in the budget constraint. An illustration is found in Figure 3.1, which for ease of presentation has been drawn for a single individual. The individual receives an unemployment benefit equal to b. Other nonlabour income equals I. The budget constraint consists of the point B and the line AC. So, if the individual works non-zero hours he looses his benefit. This is a simplification of reality, because if an unemployed person starts earning money but his total income is below the official poverty line, his income is supplemented with welfare benefits. In Figure 3.1 two situations are shown that can occur in the case of nonconvex budget sets. In Figure 3.1.B the individual under consideration prefers to work h hours, since working h hours without receiving unemployment benefit b yields a higher utility (U_1) than not working and receiving unemployment benefit equal to b (U_0), whereas in Figure 3.1.A the individual prefers not to work. By means of the direct utility function the point of highest utility is computed.



A. Voluntary unemployment



B. Involuntary unemployment

Figure 3.1 Nonconvex budget sets

Thus hours worked are explained by our model as follows: Let U_{1k} be the utility of working h_k hours without unemployment compensation, and let U_{0k} be the utility if the individual doesn't work, but receives unemployment benefit b_k. U_{0k} and U_{1k} can be calculated directly by using the utility function given in equation (2.3). Assuming absence of optimization errors, the choice to work or not (or rather to look for a job or not) for an individual receiving unemployment benefit b_k, is determined by $U_{1k}-U_{0k}$. It is important to note that, contrary to what is common in search theory, the receipt of an unemployment benefit is assumed to be exogenous. As a result the model can be written as follows

For individuals who do not receive any benefit

$$
\begin{aligned}
h_k^o &= h_k + \varepsilon_{hk} && \text{if } h_k + \varepsilon_{hk} \geq 0 \\
&= 0 && \text{if } h_k + \varepsilon_{hk} < 0
\end{aligned}
\tag{3.1}
$$

For individuals who receive a benefit

$$
\begin{aligned}
h_k^o &= h_k + \varepsilon_{hk} && \text{if } h_k+\varepsilon_{hk}\geq 0 \text{ and } U_{1k}-U_{0k}+v_k\geq 0 \\
&= 0 && \text{if } h_k+\varepsilon_{hk}<0 \text{ or } U_{1k}-U_{0k}+v_k<0
\end{aligned}
\tag{3.2}
$$

where $$h_k = \delta_k + \gamma w_k + \beta[I_k + \vartheta + \delta_k w_k + 1/2\gamma w_k^2] \tag{3.3}$$

$$\delta_k = \delta_0 + \delta_1 f_k \tag{3.4}$$

$$v_k \sim N(0, \sigma_v^2) \tag{3.5}$$

$$\varepsilon_{hk} \sim N(0, \sigma_h^2) \tag{3.6}$$

The additive error term v_k can only represent an optimization error. U_{0k} and U_{1k} can be calculated by inserting respectively $(0, I_k+b_k)$ and $(h_k, I_k+w_k h_k)$ for (h_k, y_k) into equations (2.3) and (2.4). Note that h_k is the number of hours individual k would like to work if he didn't receive any benefit and b_k is the amount of unemployment benefit.

The utility of working an optimal number of hours without benefits, U^i_{1k} (i=f,m), and the utility U^i_{0k} (i=f,m) of not working and receiving benefits can be computed for two-adult-households in a similar way by using equations (2.14)-(2.16). The model for these individuals is now written as follows

For individuals who do not receive any benefit

$$\begin{aligned} h^o_{mk} &= h_{mk} + \varepsilon_{mk} && \text{if } h_{mk}+\varepsilon_{mk} \geq 0 \\ &= 0 && \text{if } h_{mk}+\varepsilon_{mk} < 0 \end{aligned} \tag{3.7}$$

$$\begin{aligned} h^o_{fk} &= h_{fk} + \varepsilon_{fk} && \text{if } h_{fk}+\varepsilon_{fk} \geq 0 \\ &= 0 && \text{if } h_{fk}+\varepsilon_{fk} < 0 \end{aligned} \tag{3.8}$$

For individuals who receive a benefit

$$\begin{aligned} h^o_{mk} &= h_{mk} + \varepsilon_{mk} && \text{if } h_{mk}+\varepsilon_{mk} \geq 0 \text{ and } U^m_{1k}-U^m_{0k}+v_{mk} \geq 0 \\ &= 0 && \text{if } h_{mk}+\varepsilon_{mk} < 0 \text{ or } U^m_{1k}-U^m_{0k}+v_{mk} < 0 \end{aligned} \tag{3.9}$$

$$\begin{aligned} h^o_{fk} &= h_{fk} + \varepsilon_{fk} && \text{if } h_{fk}+\varepsilon_{fk} \geq 0 \text{ and } U^f_{1k}-U^f_{0k}+v_{fk} \geq 0 \\ &= 0 && \text{if } h_{fk}+\varepsilon_{fk} < 0 \text{ or } U^f_{1k}-U^f_{0k}+v_{fk} < 0 \end{aligned} \tag{3.10}$$

where $$\begin{bmatrix} \varepsilon_{mk} \\ \varepsilon_{fk} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_m & \rho\sigma_m\sigma_f \\ \rho\sigma_m\sigma_f & \sigma^2_f \end{bmatrix}\right] \tag{3.11}$$

$$\begin{bmatrix} v_{mk} \\ v_{fk} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{vm} & 0 \\ 0 & \sigma^2_{vf} \end{bmatrix}\right] \tag{3.12}$$

and v_{mk} is assumed uncorrelated with ε_{mk} and ε_{fk} and so is v_{fk} with ε_{fk} and ε_{mk}, h_{mk} and h_{fk} are given by equations (2.10)-(2.12), and U_{1k} and U_{0k} can be calculated from (2.13)-(2.16). Like in the model for singles, the additive error terms v_{mk} and v_{fk} can only represent optimization errors on the part of the individual.

3.3 The use of preferred hours

Suppose that there are institutional constraints on the number of hours one can work; say one can only work zero or fourty hours a week. This situation is sketched in Figure 3.2.

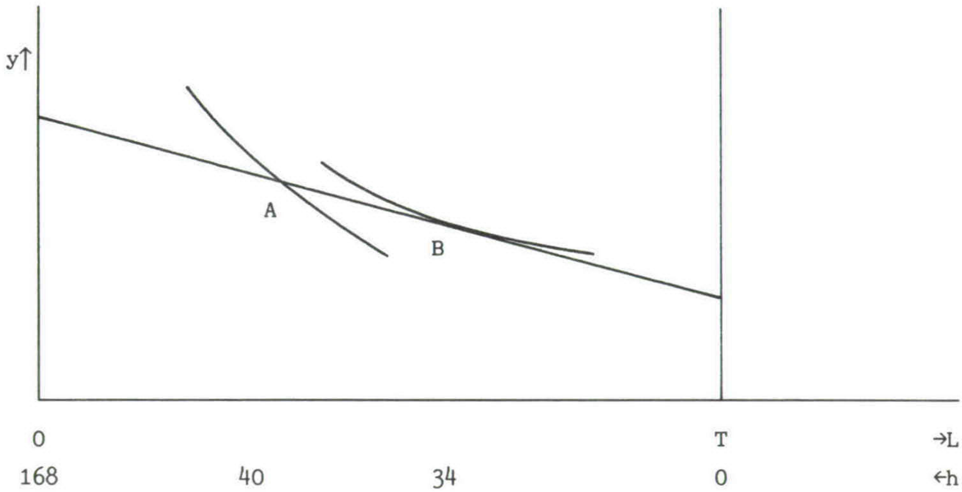


Figure 3.2 Institutional constraints

Point B would represent a utility maximum, but actually point A is observed. The latter point does not convey information about the utility function, because it does not represent a point of tangency of an indifference curve with the budget curve. Yet, in most empirical work it is assumed that observed hours are also preferred hours.

A rather straightforward way to find out whether observed behaviour represents a utility maximum or a constraint of some kind (apart from the budget constraint) is to ask respondents in a survey directly how many hours they would like to work. In the surveys that serve as data sources for the empirical work in this study adults in each participating household are asked the following question

"Suppose you could freely choose the number of hours you work per week. How many hours would you like to work in your present job, if you could choose them yourself and if you would earn on average the same amount of money per hour as you do at the moment. If you choose fewer hours of work, you choose for less income. And more hours of work means more income. Assume that the number of hours of other members of the household, if any, do not change".

We will refer to this question as the "preferred hours question" and the answer to the question as the number of preferred hours of the respondent. Figure 3.3 illustrates how the answer to this question should be interpreted within the model.

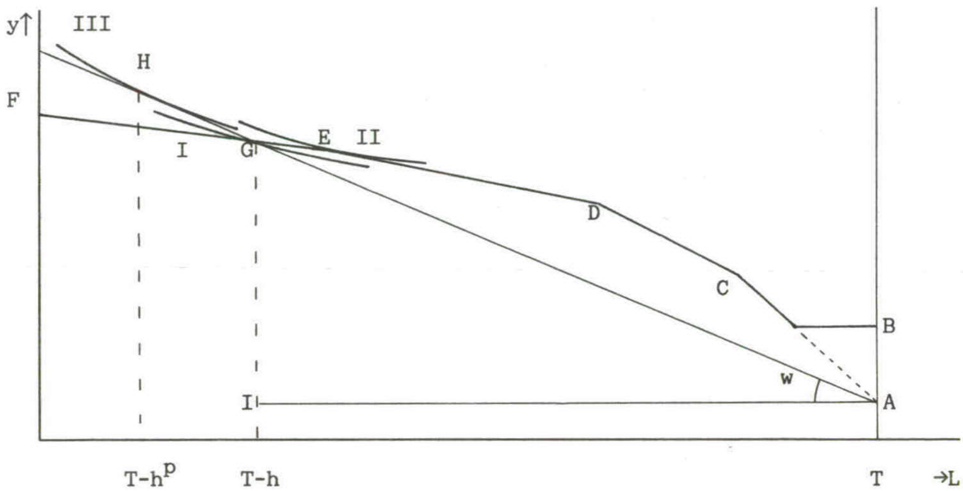


Figure 3.3 Interpretation of the response to the preferred hours question

The budget constraint faced by the individual is BCDEF. Non-labour income equals AT. The optimal number of hours for the individual is E, since indifference curve II is the highest attainable one given the budget constraint. Due to institutional constraints the actual number of hours worked by the individual deviates from the optimum. The actual number of hours is h and the indifference curve I passing through G represents a lower utility level than indifference curve II. In the actual situation labour income equals GI. Now consider the preferred hours question. In this question the respondent is invited to keep the average after tax wage rate constant. In other words, the respondent is offered a linear budget constraint AGH and is asked which number of hours is optimal. The answer is h^p. Obviously the point H does give us information on the respondent's utility function because the indifference curve is tangent to the "virtual" budget constraint and both the wage rate w and unearned income AT are known. Thus by using information on preferred hours h_k^p, wages w_k and non-labour incomes I_k for individual k in some sample we can estimate the parameters in (2.1) and thereby measure utility functions of consumption and hours worked. The model we have estimated is described by equations (3.1)-(3.6), except that h_k^o, the observed number of hours individual k actual works, is now substituted by h_k^p, the observed number of hours an individual prefers to work.

There is an essential complication in the two-adult household case which was not present in the one-adult household case. Recall the last sentence of the preferred hours question which suggests to the respondent to keep actual hours of other family members constant. In terms of equations (2.10)-(2.12) this means for example that if a male respondent says he would like to work h_{mk} hours he assumes that the number of hours worked by his wife remains at its actual level. The actual number of hours worked by the female need not correspond to a household utility maximum (e.g., due to institutional constraints). In more technical terms, the female may be rationed at her actual number of hours. Denote her actual number of hours by h_{fk}^o. Then we assume that the male response to the preferred hours question is the result of maximization of $U(h_{mk}, h_{fk}, y_k)$ subject to

$$h_{fk} = h_{fk}^o \tag{3.13}$$

$$y_k = w_{mk}h_{mk} + w_{fk}h_{fk} + I_k \tag{3.14}$$

where h_{fk}^o is the actual number of hours worked by the female (see also Neary and Roberts (1980)). Analogously, a female's response to the preferred hours question is assumed to be the result of maximizing a household utility function with h_{mk} rationed at the actual number of hours h^{o}_{mk} hours and vice versa.

Let h^{r}_{mk} and h^{r}_{fk} be the preferred number of hours of the male and the female in family k, given that the partner is rationed at his or her actual number of hours. The values of h^{r}_{mk} and h^{r}_{fk} are then generated by rationed versions of (2.10) - (2.12) as follows (see Kapteyn, Kooreman and Van Soest (1989))

$$h^{r}_{mk} = \delta_{mk} + \gamma_{m} w_{mk} + \alpha \bar{w}_{fk} + \beta_{m} \bar{I}^{f*}_{k} \tag{3.15}$$

$$h^{r}_{fk} = \delta_{fk} + \gamma_{f} w_{fk} + \alpha \bar{w}_{mk} + \beta_{f} \bar{I}^{m*}_{k} \tag{3.16}$$

$$\begin{aligned}\bar{I}^{f*}_{k} &= I_{k} + \vartheta + \delta_{mk} w_{mk} + \delta_{fk} \bar{w}_{fk} + \tfrac{1}{2}(\gamma_{m} w^{2}_{mk} + \gamma_{f} \bar{w}^{2}_{fk}) + \\ &\quad + \alpha w_{mk} \bar{w}_{fk} + h^{o}_{fk}(w_{fk} - \bar{w}_{fk})\end{aligned} \tag{3.17}$$

$$\begin{aligned}\bar{I}^{m*}_{k} &= I_{k} + \vartheta + \delta_{mk} \bar{w}_{mk} + \delta_{fk} w_{fk} + \tfrac{1}{2}(\gamma_{m} \bar{w}^{2}_{mk} + \gamma_{f} w^{2}_{fk}) + \\ &\quad + \alpha \bar{w}_{mk} w_{fk} + h^{o}_{mk}(w_{mk} - \bar{w}_{mk})\end{aligned} \tag{3.18}$$

$$h^{o}_{mk} = \delta_{mk} + \gamma_{m} \bar{w}_{mk} + \alpha w_{fk} + \beta_{m} \bar{I}^{m*}_{k} \tag{3.19}$$

$$h^{o}_{fk} = \delta_{fk} + \gamma_{f} \bar{w}_{fk} + \alpha w_{mk} + \beta_{f} \bar{I}^{f*}_{k} \tag{3.20}$$

Equations (3.19) and (3.20) define the shadow wages $\bar{w}_{mk}$ and $\bar{w}_{fk}$.

To make this an estimable model, a stochastic specification is added to the model. For individuals who do not receive a benefit this yields the following explanation of preferred hours

$$\begin{aligned} h^{p}_{mk} &= h^{r}_{mk} + \varepsilon_{mk} && \text{if } h^{r}_{mk} + \varepsilon_{mk} \geq 0 \\ &= 0 && \text{if } h^{r}_{mk} + \varepsilon_{mk} < 0 \end{aligned} \tag{3.21}$$

$$\begin{aligned} h^{p}_{fk} &= h^{r}_{fk} + \varepsilon_{fk} && \text{if } h^{r}_{fk} + \varepsilon_{fk} \geq 0 \\ &= 0 && \text{if } h^{r}_{fk} + \varepsilon_{fk} < 0 \end{aligned} \tag{3.22}$$

For individuals who receive unemployment benefit

$$h^p_{mk} = h^r_{mk} + \varepsilon_{mk} \text{ if } h^r_{mk}+\varepsilon_{mk} \geq 0 \text{ and } U^m_{1k}-U^m_{0k}+v_{mk} \geq 0$$
$$= 0 \qquad \text{if } h^r_{mk}+\varepsilon_{mk}<0 \text{ or } U^m_{1k}-U^m_{0k}+v_{mk}<0 \tag{3.23}$$

$$h^p_{fk} = h^r_{fk} + \varepsilon_{fk} \text{ if } h^r_{fk}+\varepsilon_{fk} \geq 0 \text{ and } U^f_{1k}-U^f_{0k}+v_{fk} \geq 0$$
$$= 0 \qquad \text{if } h^r_{fk}+\varepsilon_{fk}<0 \text{ or } U^f_{1k}-U^f_{0k}+v_{fk}<0 \tag{3.24}$$

where $$\begin{pmatrix}\varepsilon_{mk}\\ \varepsilon_{fk}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2_m & \rho\sigma_m\sigma_f\\ \rho\sigma_m\sigma_f & \sigma^2_f\end{pmatrix}\right) \tag{3.25}$$

$$\begin{pmatrix}v_{mk}\\ v_{fk}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma^2_{vm} & 0\\ 0 & \sigma^2_{vf}\end{pmatrix}\right) \tag{3.26}$$

and v_{mk} is assumed uncorrelated with ε_{mk} and ε_{fk} and so is v_{fk} with ε_{fk} and ε_{mk}, h^p_{mk} and h^p_{fk} are observed preferred hours of man and woman, h^r_{mk} and h^r_{fk} are given in equations (3.15)-(3.20). The additive error terms in the hours equation, ε_{mk} and ε_{fk}, can represent measurement errors or optimization errors. But they cannot represent random preferences: Because if part of ϑ is random and let us say normally distributed, then the error term in the hours equation would be a nonlinear function of the normally distributed random term in ϑ and as a consequence it cannot be normally distributed. This is due to the fact that the shadow wages are nonlinear functions of the hours variables. The v_{mk} and v_{fk} can only represent optimization errors.

One more complication with respect to the preferred hours question arises. Nonworking respondents are asked a different question, namely whether they are looking for a job or not. If, for example, an individual reports not to be looking for a job this is taken as evidence that the utility of not working is higher than the utility of working.

3.4 Estimation results

- Actual hours, recipients of an unemployment benefit included

The model specification for the standard model in which the nonconvexity of the budget constraint is taken into account, are given by equations (3.1)-(3.6) for singles and by equations (3.7)-(3.12) for two-adult households. The endogenous variable is actual hours worked by the individual. The likelihood function is given in Appendix 3A. Note that the results are based on samples which include individuals receiving an unemployment benefit (in contrast with the samples used in Chapter 2). The sample composition is given in Table 3.1 (compare Tables 2.1 and 2.4).

In Table 3.2 the estimation results for singles are given. Comparing Table 3.2 and 2.6 reveals that including individuals who receive an unemployment benefit does not change the estimated parameter values very much. The parameter σ_v represents the standard deviation of the random component in the utility difference between working an optimal number of hours without unemployment benefits and working zero hours with unemployment benefits. Its estimated value for males is very inaccurately determined, mainly because in the sample there is only a small number of individuals who have to make this utility comparison (cf. Table 3.1). For females the value of σ_v is more accurately determined. In general the same conclusions of no large changes in the estimated parameter values hold for two-adult households (Table 3.3). Only for the OSA-data set the insignificant female wage coefficient changes from 2 to 0.2. All standard errors of the random components in the utility differences (σ_v's) are inaccurately determined and all are large in value, except for single females. In Table 3.4 wage elasticities are shown. All own wage elasticities for males are positive. They have increased slightly compared with the elasticities shown in Table 2.8. The female own wage elasticities are still much higher than the male ones. For the OSA-sample the numbers have decreased a bit in comparison with the ones shown in Chapter 2. One can conclude that including behaviour of individuals who receive a benefit into the standard model hardly changes the estimation results of the model (in Chapter 7 we will present a formal test).

Table 3.1 Sample composition[a)]

OSA	male	h>0	h=0				
			seek=0		seek=1		
female			b=0	b>0	b=0	b>0	
h>0		314	2	1	4	9	330 (131)
h=0, seek=0,	b=0	449	0	0	0	25	474 (15)
	b>0	3	0	0	0	1	4 (21)
h=0, seek=1,	b=0	26	0	0	0	4	30 (4)
	b>0	6	0	0	0	2	8 (27)
all		798	2	1	4	41	846 (341=198+
		(112)	(0)	(2)	(1)	(28)	143)

SEP	male	h>0	h=0				
			seek=0		seek=1		
female			b=0	b>0	b=0	b>0	
h>0		296	5	3	0	1	305 (77)
h=0, seek=0,	b=0	567	20	23	0	10	620 (49)
	b>0	6	1	0	0	0	7 (66)
h=0, seek=1,	b=0	19	0	0	0	1	20 (0)
	b>0	4	0	0	0	0	4 (12)
all		892	26	26	0	12	956=(371=204+
		(113)	(9)	(23)	(1)	(21)	167

a) The numbers in parentheses refer to singles. Individuals who work (h>0) and receive an unemployment benefit (b>0) have been excluded. For definition of the variable "seek" see Table 2.2.

Table 3.2 Estimation results for the standard model, including individuals who receive unemployment benefits, one-adult households[a)]

OSA

	males		females	
δ_0	36	(6)	14	(3)
β	-0.000001	(u.b.)	-0.038	(0.005)
δ_1	0.4	(6)	-8	(4)
γ	0.2	(0.5)	2.0	(0.2)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	11.7	(0.5)	15.5	(1.0)
σ_v	1932	(4257)	155	(65)
log likelihood	-438.1		-568.0	

SEP

	males		females	
δ_0	14	(10)	-25	(8)
β	-0.009	(0.003)	-0.069	(0.008)
δ_1	4.8	(3.8)	-1	(15)
γ	0.8	(0.6)	2.5	(0.4)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	16.0	(1.0)	22.6	(1.9)
σ_v	32883	(337671)	136	(32)
log likelihood	-517.0		-414.4	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.3 Estimation results for the standard model, including individuals who receive unemployment benefits, two-adult households[a)]

	OSA		SEP	
α	-0.2	(0.5)	0.4	(0.1)
β_m	-0.003	(0.002)	-0.023	(0.001)
β_f	-0.02	(0.006)	-0.000001	(u.b.)
γ_m	0.2	(0.2)	0.01	(ℓ.b.)
γ_f	0.2	(4.9)	4.1	(0.6)
δ_{m1}	-0.4	(1.5)	-3	(2)
δ_{f1}	-62	(8)	-41	(4)
δ_{m0}	31	(15)	-17	(5)
δ_{f0}	-45	(128)	-12	(7)
ϑ	-3632	(5361)	-2211	(201)
σ_m	6.91	(0.08)	11.5	(0.2)
σ_f	24.3	(1.6)	25.4	(1.8)
ρ	-0.04	(0.04)	0.14	(0.04)
σ_{mv}	73846	(25337)	6.10^6	(1.10^7)
σ_{fv}	4.10^6	(4.10^8)	2470	(14283)
log likelihood	-4499.3		-5299.3	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.4 Wage elasticities for the standard model, including individuals who receive unemployment benefits[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m} \frac{w_m}{h_m}$	0.08	0.02	0.20	0.11
	(0.16)	(0.06)	(0.21)	(0.03)
$\frac{\partial h_f}{\partial w_f} \frac{w_f}{h_f}$	0.23	1.83	0.94	5.02
	(0.08)	(1.28)	(0.50)	(5.15)

a) Standard errors in parentheses

- Preferred hours

The model for singles which has been estimated is given by equations (2.7)-(2.9), except that the dependent variable h_k^o has been replaced by h_k^p (observed preferred hours of work). Equations (2.10)-(2.12) and (2.18)-(2.22) display the model for two-adult households with h_{mk}^o and h_{fk}^o replaced by h_{mk}^p and h_{fk}^p . In Tables 3.5 and 3.6 the estimation results are shown for the model in which preferred hours is used as the dependent variable and recipients of benefits are not included. By comparing Tables 2.6 and 2.7 with Tables 3.5 and 3.6 it can be seen what effect the change in the dependent variable has on the parameter values. We would have expected that the wage and income coefficients are lower in the actual hours version than in the preferred hours version. The obvious explanation for this is that actual hours are partly determined by institutional constraints. This is only true for single males (OSA). In most other cases the wage and income coefficients are slightly lower in the preferred hours version, but not significantly so. In Table 3.7 the wage elasticities corresponding with this model are presented. In some cases (two-adult-households (OSA), one-adult-households (SEP)) the own male wage elasticity becomes negative. Almost all wage elasticities, both for men and for women, have decreased compared with the actual hours version of the model. The standard errors of the elasticities decreased a little. In Tables 3.8, 3.9 and 3.10 estimation results are presented for the model with preferred hours as the endogenous variable and where recipients of benefits are included. This yields similar results as compared to the model in which recipients of benefits are excluded. (Compare Tables 3.5 and 3.6 with Tables 3.8 and 3.9.) This is in correspondence with the results earlier mentioned in this section, namely that including recipients of benefits does not change the estimation results very much.

In Table 3.10 the rationed version of the above mentioned model is shown (equations 3.15-3.20). The estimated parameter values differ quite a bit from the ones shown in Table 3.9. Most striking is the fact that the linear wage coefficients (γ) and the income coefficients in the SEP-sample have increased. From a theoretical point of view this is the model we should use. The numbers in Table 3.11, in which the wage elasticities are presented, show in general somewhat lower elasticities than the model which has actual hours as the dependent variable (Table 3.4). This is also clear from the labour supply curves presented in Figure 3.4 and 3.5. From Figures 3.6 and 3.7 it is most interesting to see that there is more variation in the preferred hours than in the actual hours distribution. Compare Figures 2.3 and 2.4. The general pattern of the figures, however, is the same as the one we have seen in Chapter 2. The simulated hours distribution is not able to reproduce the spikes in the preferred hours distribution. The model predicts that all men prefer to work about 40 hours per week. The percentage of nonworkers is largely underestimated, especially in the OSA-sample.

Table 3.5 Estimation results for the standard model, with preferred hours as the dependent variable, recipients of benefits not included, one-adult households[a)]

OSA

	males		females	
δ_0	16	(17)	9	(8)
β	-0.01	(0.01)	-0.034	(0.004)
δ_1	4	(4)	-6.0	(4.0)
γ	0.5	(0.3)	1.9	(0.5)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	8.3	(0.6)	13.6	(0.9)
log likelihood				
	-392.9		-526.6	

SEP

	males		females	
δ_0	22	(10)	-20	(9)
β	-0.008	(0.003)	-0.061	(0.007)
δ_1	3	(3)	9	(11)
γ	0.1	(0.6)	2.4	(1.9)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	16.3	(0.9)	20.3	(1.8)
log likelihood				
	-488.1		-376.0	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.6 Estimation results for the standard model, with preferred hours as the dependent variable, recipients of benefits not included, two-adult households[a)]

	OSA		SEP	
α	-0.4	(1.5)	-0.1	(0.2)
β_m	-0.002	(0.002)	-0.022	(0.0009)
β_f	-0.016	(0.005)	-0.000001	(u.b.)
γ_m	0.01	(ℓ.b.)	0.01	(ℓ.b.)
γ_f	2	(11)	3.8	(0.6)
δ_{m1}	1.2	(1.0)	-5.2	(1.9)
δ_{f1}	-48	(6)	-38	(4)
δ_{m0}	37	(70)	-2	(5)
δ_{f0}	0.1	(535)	-4	(6)
ϑ	-2129	(33986)	-1814	(149)
σ_m	6.3	(0.1)	9.3	(0.1)
σ_f	21.2	(1.4)	23.8	(1.4)
ρ	-0.12	(0.04)	0.02	(0.04)
log likelihood				
	-4271.3		-4955.7	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.7 Wage elasticities for the standard model, with preferred hours as the dependent variable, recipients of benefits not included[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.09	-0.02	-0.04	0.08
	(0.12)	(0.06)	(0.20)	(0.04)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.34	2.00	0.82	4.95
	(0.21)	(2.66)	(0.31)	(4.81)

a) Standard errors in parentheses

Table 3.8 Estimation results for the standard model, with preferred hours as the dependent variable, recipients of benefits included, one-adult households[a)]

OSA

	males		females	
δ_0	17	(18)	13.5	(3.3)
β	-0.01	(0.01)	-0.038	(0.005)
δ_1	4	(4)	-8.3	(4.3)
γ	0.5	(0.3)	2.00	(0.22)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	8.3	(0.6)	15.5	(1.0)
σ_v	971	(1250)	155	(65)
log likelihood	-400.3		-568.1	

SEP

	males		females	
δ_0	19	(10)	-21	(9)
β	-0.009	(0.003)	-0.06	(0.01)
δ_1	4	(3)	4	(11)
γ	0.3	(0.6)	2.4	(0.4)
ϑ	-1650	(fixed)	-355	(fixed)
σ_h	16.3	(0.9)	21	(2)
σ_v	52118	(136502)	145	(34)
log likelihood	-518.8		-409.9	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.9 Estimation results for the standard model, with preferred hours as the dependent variable, recipients of benefits included, two-adult households[a)]

	OSA		SEP	
α	-0.3	(0.3)	0.194	(0.006)
β_m	-0.002	(0.002)	-0.018	(0.001)
β_f	-0.018	(0.006)	-0.000001	(u.b.)
γ_m	0.1	(0.2)	0.01	(ℓ.b.)
γ_f	2.7	(2.6)	3.8	(0.2)
δ_{m1}	1.0	(1.0)	-2.0	(0.1)
δ_{f1}	-48	(6)	-39	(3)
δ_{m0}	40	(15)	-0.4	(0.1)
δ_{f0}	30	(105)	-7.9	(1.6)
ϑ	-361	(5993)	-1814	(109)
σ_m	6.3	(0.1)	11.3	(0.2)
σ_f	21.6	(1.4)	24.6	(1.3)
ρ	-0.12	(0.04)	0.11	(0.04)
σ_{vm}	6.10^6	(8.10^7)	716	(110)
σ_{vf}	4.10^6	(7.10^7)	1000	(3064)
log likelihood	-4392.6		-5264.1	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.10 Estimation results for the standard model, with preferred hours as the dependent variable, recipients of benefits included, rationed version, two-adult households[a)]

	OSA		SEP	
α	0.9	(0.4)	1.7	(0.3)
β_m	-0.0014	(0.0004)	-0.023	(0.002)
β_f	-0.0094	(0.0002)	-0.012	(0.004)
γ_m	0.28	(0.14)	1.48	(0.34)
γ_f	8.7	(3.6)	4.20	(0.63)
δ_{m1}	1.7	(0.7)	-22.8	(8.4)
δ_{f1}	-45.4	(4.4)	-47.8	(8.5)
δ_{m0}	120	(33)	13.6	(7.1)
δ_{f0}	612	(272)	-37.5	(11.8)
ϑ	62565	(29264)	713	(358)
σ_m	6.20	(0.07)	12.8	(0.2)
σ_f	21.9	(1.4)	24.9	(1.1)
ρ	-0.041	(0.058)	-0.723	(0.019)
σ_{vm}	2616	(1865)	1661	(593)
σ_{vf}	3209	(27062)	3882	(21185)
log likelihood	-4386.7		-5220.4	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.

Table 3.11 Wage elasticities for the standard model, with preferred hours as the dependent variable, recipients of benefits included[a)]

	OSA			SEP		
	one-adult	two-adult		one-adult	two-adult households	
		no rat.	with rat.		no rat.	with rat.
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.09	-0.01	0.02	0.03	0.01	0.03
	(0.12)	(0.06)	(0.00)	(0.20)	(0.00)	(0.08)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.23	1.83	1.74	0.89	5.05	2.17
	(0.08)	(1.38)	(1.29)	(0.50)	(2.89)	(4.21)

a) Standard errors in parentheses.

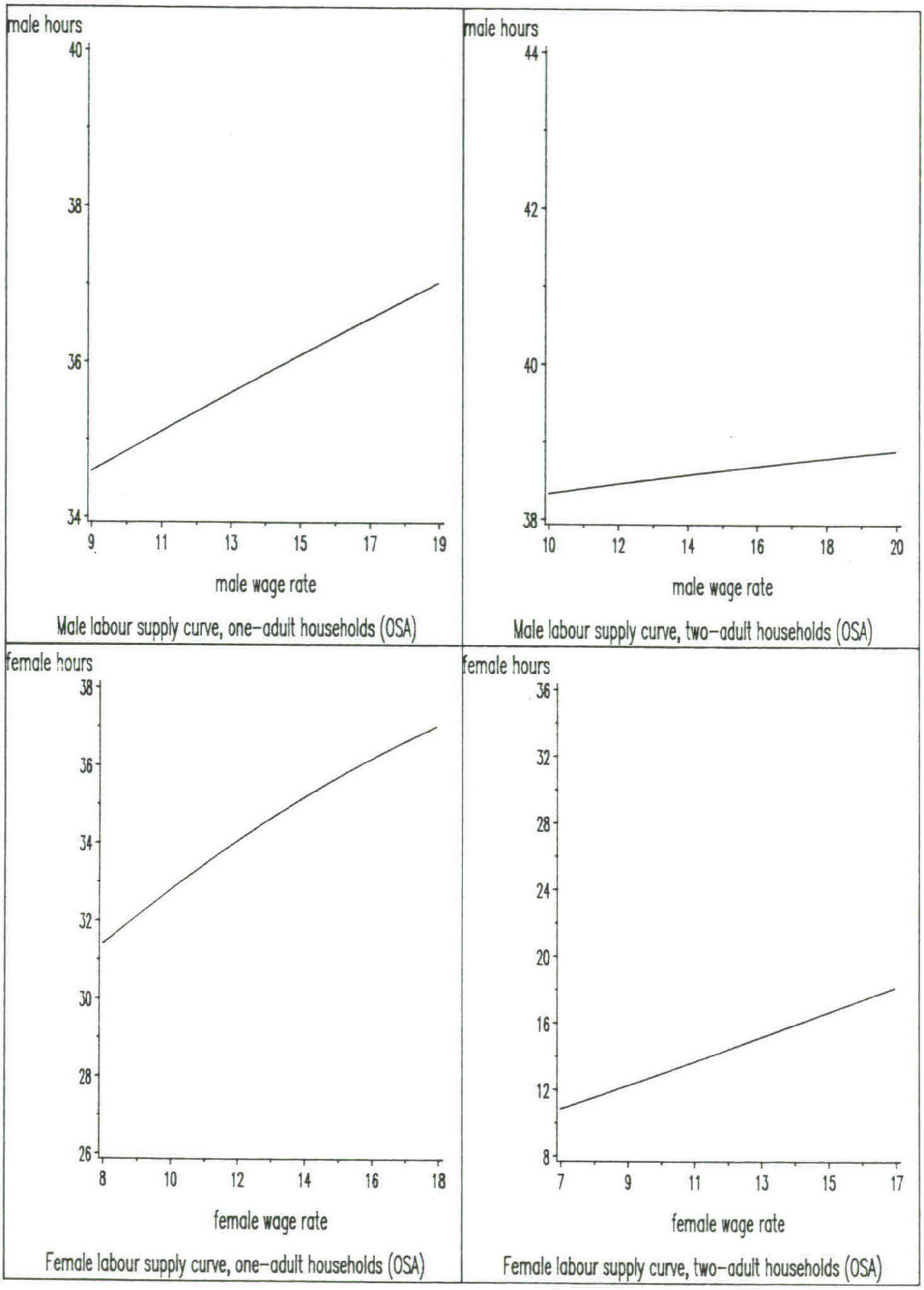


Figure 3.4 Labour supply curves (OSA)

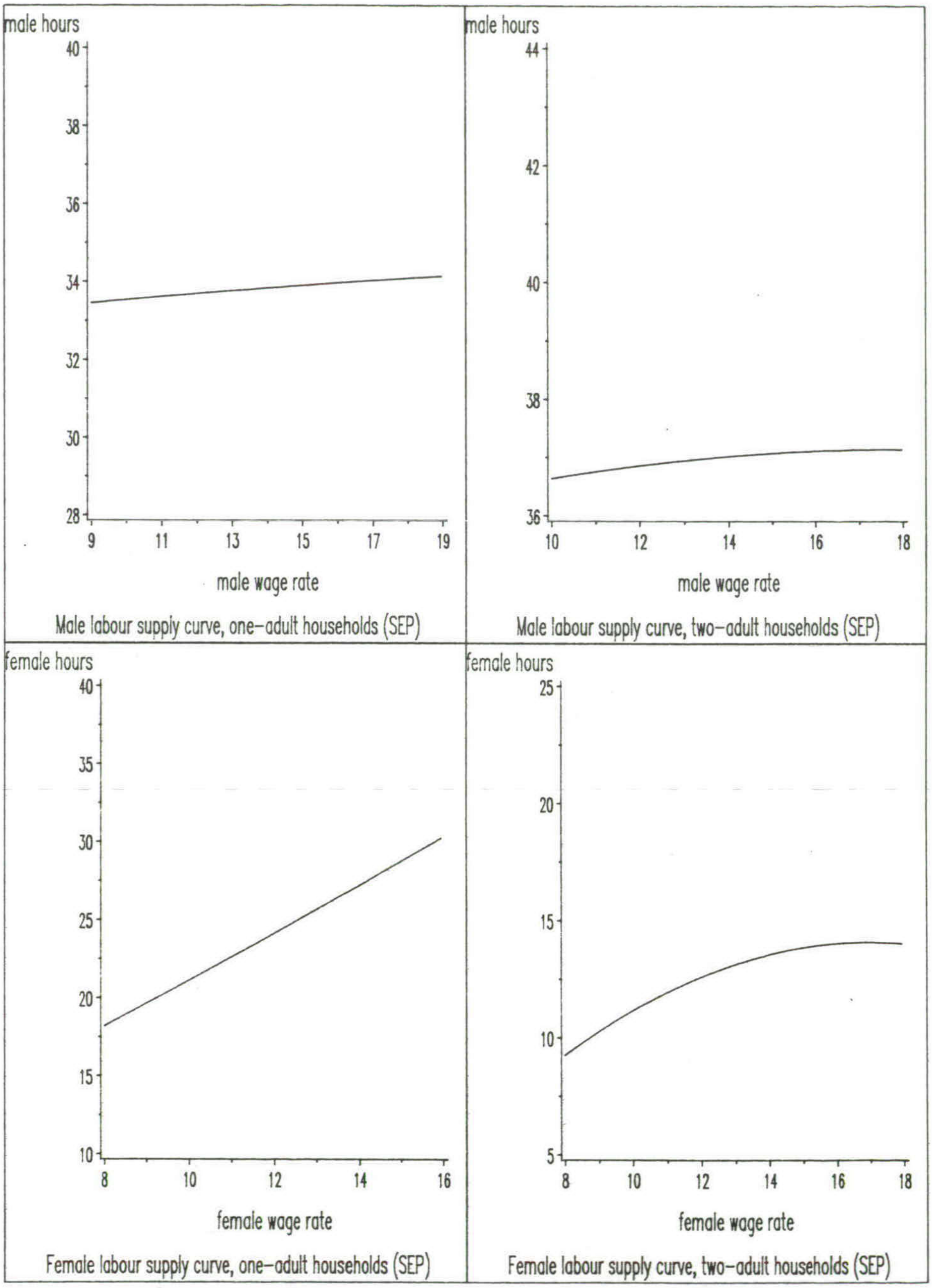


Figure 3.5 <u>Labour supply curves (SEP)</u>

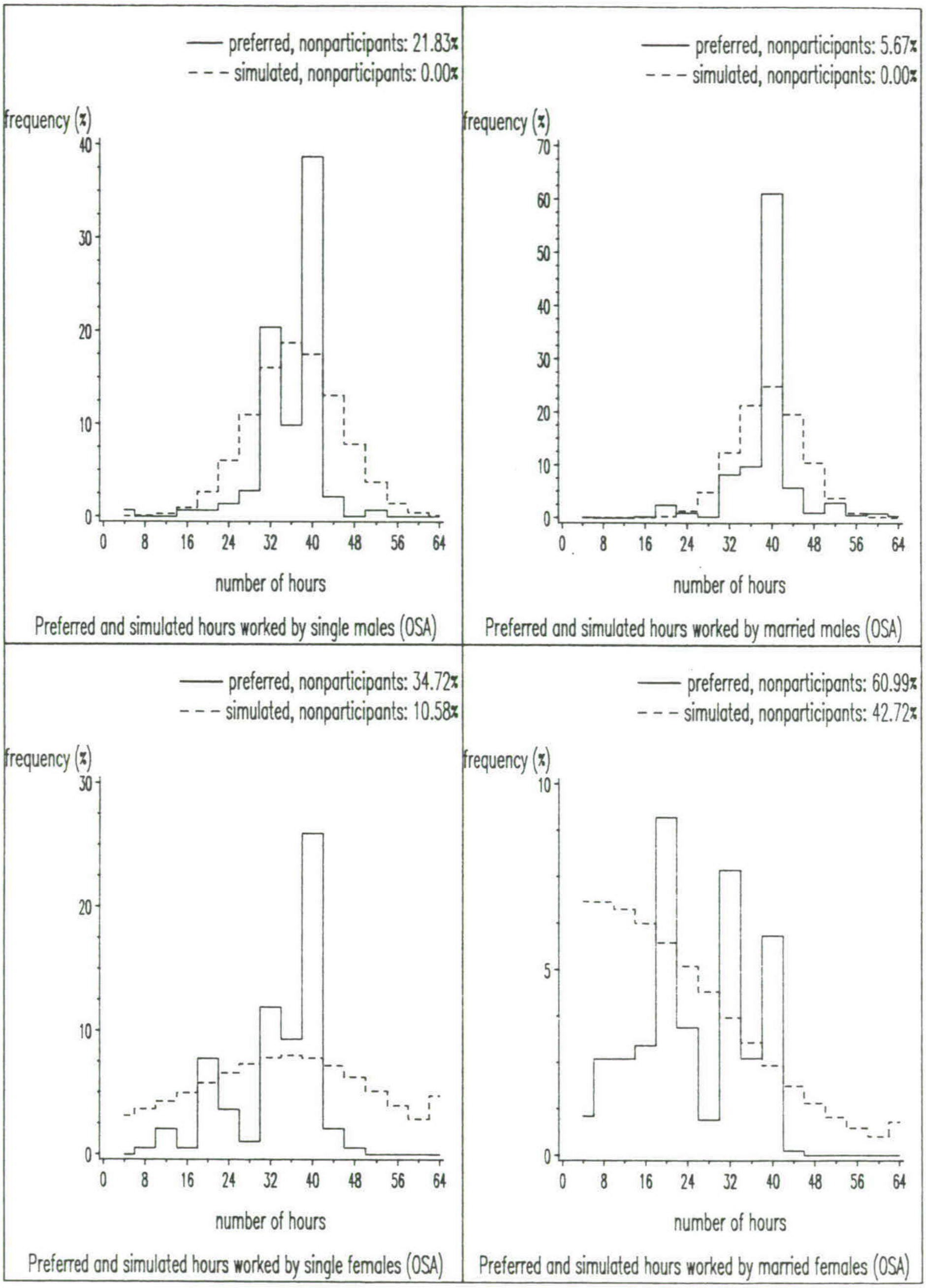


Figure 3.6 <u>Hours distributions (OSA)</u>

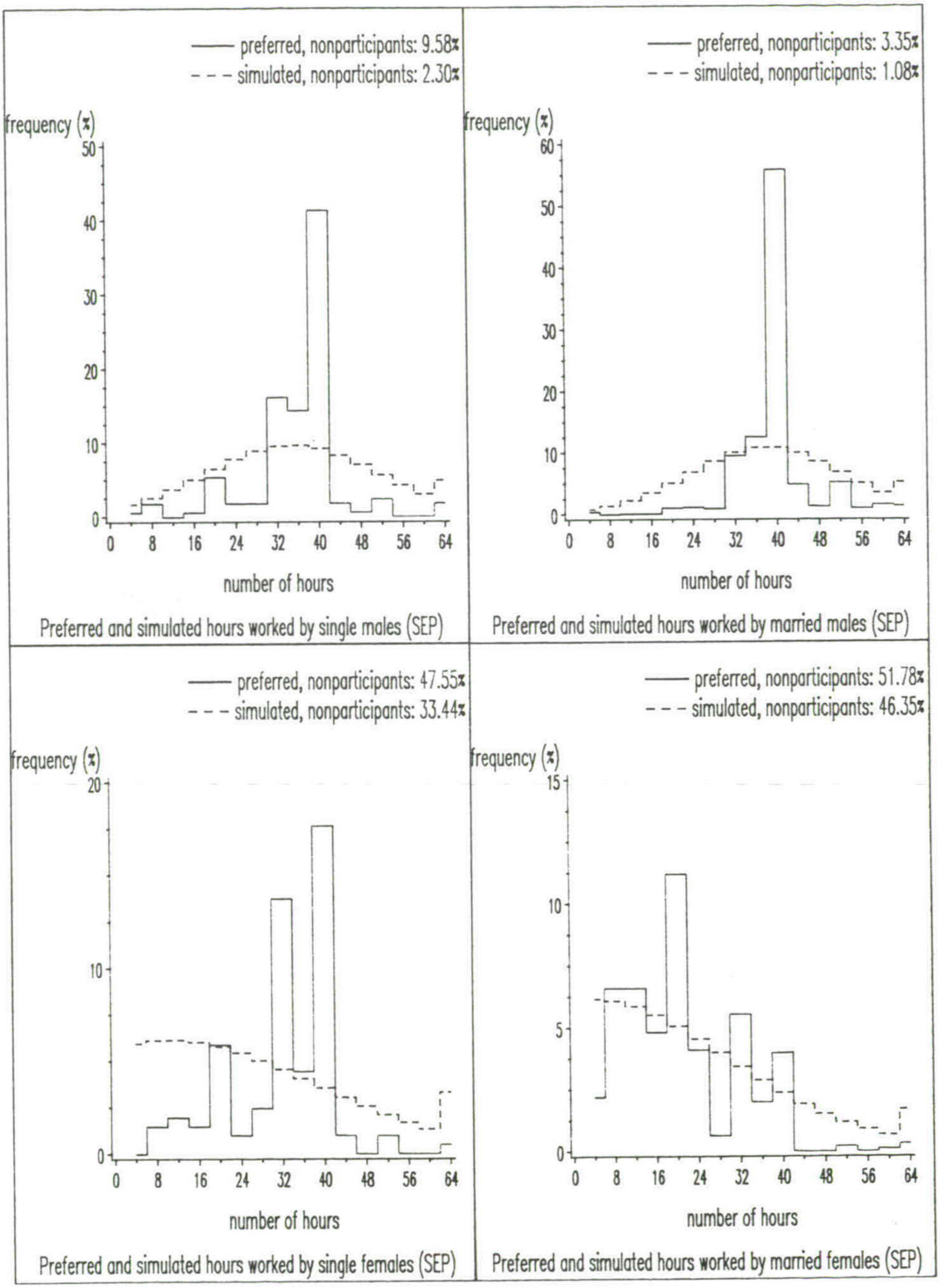


Figure 3.7 <u>Hours distributions (SEP)</u>

3.5 Concluding remarks

In Section 2 of this chapter a first step towards a more realistic model is made by modelling the social security and welfare system which lead to a nonconvex budget set. We find that taking into account the non-convexity of the budget constraint hardly changes the estimation results.

Since we believe that actual hours worked is not the appropriate endogenous variable in a labour supply model, we introduce in Section 3 the preferred hours variable.

Contrary to our ex-ante expectations, both the income coefficients and the wage elasticities are lower in the preferred hours version than in the actual hours version (see Section 3.4).

Appendix 3A. Likelihood contributions

In this appendix the likelihood functions for both the one-adult household the two-adult household model, described in Sections 3.2 and 3.3 will be given.

The likelihood function consists of different parts (L_k) corresponding to the different situations households face. Let Φ and φ be the standard normal distribution and density function respectively. Let $B\Phi$ and $b\varphi$ be the bivariate standard normal distribution and density functions, respectively (with correlation ρ). First we will derive the likelihood function for the models described in Section 2.

- One-adult households

The following situations are distinguished

1) individual k is working ($h_k^o>0$)

$$L_k^1 = \frac{1}{\sigma_h} \varphi\left[\frac{h_k^o - h_k}{\sigma_h} \right] \tag{3A.1}$$

2) individual k is not working and not seeking a job and does not receive an unemployment benefit ($h_k^o=0$, $seek_k=0$, $b_k=0$)

$$L_k^2 = \Phi\left[\frac{-h_k}{\sigma_h} \right] \tag{3A.2}$$

3) individual k is not working but seeking a job and does not receive an unemployment benefit ($h_k^o=0$, $seek_k=1$, $b_k=0$)

$$L_k^3 = \Phi\left[\frac{h_k}{\sigma_h} \right] \tag{3A.3}$$

4) individual k is not working and not seeking a job and receives an unemployment benefit ($h_k^o=0$, $seek_k=0$, $b_k>0$)

$$L_k^4 = 1 - \Phi\left(\frac{h_k}{\sigma_h}\right)\Phi\left[\frac{\left[U_{1k} - U_{0k}\right]}{\sigma_v}\right] \tag{3A.4}$$

5) individual k is not working but seeking a job and receives an unemployment benefit ($h_k^o=0$, $seek_k=1$, $b_k>0$

$$L_k^5 = \Phi\left(\frac{h_k}{\sigma_h}\right)\Phi\left[\frac{\left[U_{1k} - U_{0k}\right]}{\sigma_v}\right] \tag{3A.5}$$

- Two-adult households

We distinguish the following situations: (where i,j stands for m(ale) or f(emale))

1) both spouses are working and neither receives an unemployment benefit ($h_{ik}^o>0$, $h_{jk}^o>0$, $b_{ik}=0$, $b_{jk}=0$)

$$L_k^1 = \frac{1}{\sigma_i\,\sigma_j}\, b\varphi\left[\frac{h_{ik}^o - h_{ik}}{\sigma_i}, \frac{h_{jk}^o - h_{jk}}{\sigma_j}, \rho\right] \tag{3A.6}$$

2) spouse i is not working and not seeking a job, and spouse j is working and neither receives an unemployment benefit ($h_{ik}^o=0$, $seek_{ik}=0$, $h_{jk}^o>0$, $b_{ik}=0$, $b_{jk}=0$)

$$L_k^2 = \Phi\left[\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left[h_{jk}^o - h_{jk}\right]}{\sigma_j\sqrt{1-\rho^2}}\right]\frac{1}{\sigma_j}\,\varphi\left[\frac{h_{jk}^o - h_{jk}}{\sigma_j}\right] \tag{3A.7}$$

3) spouse i is not working and not seeking a job and receives an unemployment benefit, and spouse j is working ($h_{ik}^o=0$, $seek_{ik}=0$, $h_{jk}^o>0$, $b_{ik}>0$, $b_{jk}=0$)

$$L_k^3 = \left(1- \left[1-\Phi\left(\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left(h_{jk}^o - h_{jk}\right)}{\sigma_j\sqrt{1-\rho^2}} \right) \right] \Phi\left(\frac{\left(U_{1k}^i - U_{0k}^i\right)}{\sigma_{vi}} \right) \right) \frac{1}{\sigma_j}\varphi\left(\frac{h_{jk}^o - h_{jk}}{\sigma_j} \right) \tag{3A.8}$$

4) spouse i is not working but seeking a job, and spouse j is working and neither receives an unemployment benefit (h_{ik}^o =0, $seek_{ik}$=1, h_{jk}^o>0, b_{ik}=0, b_{jk}=0)

$$L_k^4 = \left[1-\Phi\left(\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left(h_{jk}^o - h_{jk}\right)}{\sigma_j\sqrt{1-\rho^2}} \right) \right] \frac{1}{\sigma_j}\varphi\left(\frac{h_{jk}^o - h_{jk}}{\sigma_j} \right) \tag{3A.9}$$

5) spouse i is not working but seeking a job and receives an unemployment benefit, and spouse j is working (h_{ik}^o =0, $seek_{ik}$=1, h_{jk}^o>0, b_{ik}>0, b_{jk}=0)

$$L_k^5 = \left[1-\Phi\left(\frac{-h_{ik}}{\sigma_i\sqrt{1-\rho^2}} - \frac{\rho\left(h_{jk}^o - h_{jk}\right)}{\sigma_j\sqrt{1-\rho^2}} \right) \right] \Phi\left(\frac{\left(U_{1k}^i - U_{0k}^i\right)}{\sigma_{vi}} \right) \frac{1}{\sigma_j}\varphi\left(\frac{h_{jk}^o - h_{jk}}{\sigma_j} \right) \tag{3A.10}$$

6) Both spouses are not working and not seeking a job and neither receives an unemployment benefit (h_{ik}^o =0, $seek_{ik}$=0, h_{jk}^o =0, $seek_{jk}$=0, b_{ik}=0, b_{jk}=0)

$$L_k^6 = B\Phi\left(\frac{-h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, \rho \right) \tag{3A.11}$$

7) Both spouses are not working and not seeking a job and only spouse i receives an unemployment benefit ($h^o_{ik}=0$, $seek_{ik}=0$, $h^o_{jk}=0$, $seek_{jk}=0$, $b_{ik}>0$, $b_{jk}=0$)

$$L^7_k = B\Phi\left[\frac{-h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, \rho\right] + B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right]\Phi\left[\frac{-\left[U^i_{1k} - U^i_{0k}\right]}{\sigma_{vi}}\right] \tag{3A.12}$$

8) Both spouses are not working, spouse i is seeking a job, spouse j is not seeking a job and neither receive an unemployment benefit ($h^o_{ik} = 0$, $seek_{ik}=1$, $h^o_{jk} = 0$, $seek_{jk}=0$, $b_{ik}=0$, $b_{jk}=0$)

$$L^8_k = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right] \tag{3A.13}$$

9) Both spouses are not working but spouse i is seeking a job and receives an unemployment benefit but spouse j is not seeking a job and does not receive an unemployment benefit ($h^o_{ik} = 0$, $seek_{ik}=1$, $h^o_{jk} = 0$, $seek_{jk}=0$, $b_{ik}>0$, $b_{jk}=0$)

$$L^9_k = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right]\Phi\left[\frac{\left[U^i_{1k} - U^i_{0k}\right]}{\sigma_{vi}}\right] \tag{3A.14}$$

10) Both spouses are not working, not seeking a job and receive an unemployment benefit ($h^o_{ik} = 0$, $seek_{ik}=0$, $h^o_{jk} = 0$, $seek_{jk}=0$, $b_{ik}>0$, $b_{jk}>0$)

$$L^{10}_k = B\Phi\left[\frac{-h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, \rho\right] + B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right]\Phi\left[\frac{-\left[U^i_{1k} - U^i_{0k}\right]}{\sigma_{vi}}\right] +$$

$$B\Phi\left[\frac{-h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, -\rho\right]\Phi\left[\frac{-\left[U_{1k}^{j} - U_{0k}^{j}\right]}{\sigma_{vj}}\right] + \tag{3A.15}$$

$$B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right]\Phi\left[\frac{-\left[U_{1k}^{i} - U_{0k}^{i}\right]}{\sigma_{vi}}\right]\Phi\left[\frac{-\left[U_{1k}^{j} - U_{0k}^{j}\right]}{\sigma_{vj}}\right]$$

11) Both spouses are not working, spouse i is seeking a job and does not receive an unemployment benefit, while spouse j is not seeking and receives a benefit ($h_{ik}^{o} = 0$, $seek_{ik}=1$, $h_{jk}^{o} = 0$, $seek_{jk}=0$, $b_{ik}=0$, $b_{jk}>0$)

$$L_k^{11} = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right] + \tag{3A.16}$$

$$B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right]\Phi\left[\frac{-\left[U_{1k}^{j} - U_{0k}^{j}\right]}{\sigma_{vj}}\right]$$

12) Both spouses are not working and receiving an unemployment benefit, spouse i is seeking a job and spouse j is not ($h_{ik}^{o} = 0$, $seek_{ik}=1$, $h_{jk}^{o} = 0$, $seek_{jk}=0$, $b_{ik}>0$, $b_{jk}>0$)

$$L_k^{12} = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{-h_{jk}}{\sigma_j}, -\rho\right]\Phi\left[\frac{\left[U_{1k}^{i} - U_{0k}^{i}\right]}{\sigma_{vi}}\right] + \tag{3A.17}$$

$$B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right]\Phi\left[\frac{\left[U_{1k}^{i} - U_{0k}^{i}\right]}{\sigma_{vi}}\right]\Phi\left[\frac{-\left[U_{1k}^{j} - U_{0k}^{j}\right]}{\sigma_{vj}}\right]$$

13) Both spouses are not working but seeking a job and neither receive an unemployment benefit ($h_{ik}^{o}=0$, $seek_{ik}=1$, $h_{jk}^{o}=0$, $seek_{jk}=1$, $b_{ik}=0$, $b_{jk}=0$)

$$L_k^{13} = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right] \tag{3A.18}$$

14) Both spouses are not working and both are seeking a job, and spouse i receives an unemployment benefit but spouse j does not ($h_{ik}^o=0$, $seek_{ik}=1$, $h_{jk}^o=0$, $seek_{jk}=1$, $b_{ik}>0$, $b_{jk}=0$)

$$L_k^{14} = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right] \Phi\left[\frac{\left[U_{1k}^i - U_{0k}^i\right]}{\sigma_{vi}}\right] \tag{3A.19}$$

15) Both spouses are not working but seeking a job and receiving an unemployment benefit ($h_{ik}^o = 0$, $seek_{ik}=1$, $h_{jk}^o = 0$, $seek_{jk}=1$, $b_{ik}>0$, $b_{jk}>0$)

$$L_k^{15} = B\Phi\left[\frac{h_{ik}}{\sigma_i}, \frac{h_{jk}}{\sigma_j}, \rho\right] \Phi\left[\frac{\left[U_{1k}^i - U_{0k}^i\right]}{\sigma_{vi}}\right]\Phi\left[\frac{\left[U_{1k}^j - U_{0k}^j\right]}{\sigma_{vj}}\right] \tag{3A.20}$$

For the likelihood contributions corresponding with the preferred hours version of the model with a convex budget constraint, described in Section 3 see Chapter 2, Section 2.5. One modification should be made with respect to these formulas, namely the endogenous variable actual hours worked (h_k^o) must be substituted by preferred hours (h_k^p). By making this same modification in equations (3A.1)-(3A.20) one obtains the preferred hours model with a nonconvex budget constraint. Finally, if h_{ik} is substituted by h_{ik}^r the likelihood function of the rationed version of the model is obtained.

4. Preference interdependence and habit formation

4.1 Introduction

It has already been stated in Chapter 2 that most likely utility functions will vary across households. For a good understanding of household labour supply it is important to identify systematic factors which cause this variation. We pay attention to two major causes of differences in utility functions across households: preference formation (in this chapter) and differences in household composition (in Chapter 2).

Preference formation refers to the phenomenon that an individual's preferences are formed, at least partly, by his past experiences and perceptions. Within the context of consumption and labour supply it may be assumed that somenone's present preferences will depend on how much he used to consume in the past and how much leisure he used to enjoy. Moreover, his preferences will probably also depend on what others do. If the individual comes from a social background where most people used to work hard one may assume that the individual's preference for leisure will be less strongly developed than in the case where most of his friends and relatives are representatives of the "leisure class".

The notion that preferences may be endogenous has gained some foothold in the literature on consumer demand equations (see e.g., Pollak (1970, 1976), Phlips (1984), Gaertner (1974), Blanciforti and Green (1983), Darrough, Pollak and Wales (1983), Alessie and Kapteyn (1985) and the references given in these papers), but hardly any in the labour supply literature. Empirically, habit formation is usually the only component of preference formation that is being modelled in consumer demand models (habit formation may then be either rational or myopic, cf. Muellbauer (1988)). In a few working papers, Alessie and Kapteyn (1985) and Kapteyn, Van de Geer, Van de Stadt and Wansbeek (1985) have also incorporated preference interdependence into empirical models of consumption following theoretical notions developed by Gaertner (1974) and Pollak (1976). In this chapter, the framework developed in the two former papers is extended to deal with household labour supply.

We will incorporate preference formation into an already fairly complex neoclassical model of household labour supply developed in Chapter

3 (See also Kapteyn, Kooreman, Van Soest (1989)). The drawback of this approach is of course its complexity.

In Section 4.2 it is extensively discussed how preference formation can be incorporated into the model developed in Chapter 3. In Section 4.3 we describe the estimation results. In Section 4.4 we will use a more sophisticated, factor analysis, model to analyse the influence of reference groups on labour supply. In Section 4.5 the estimation results of the model described in the previous section will be given. Section 4.6 concludes.

4.2 A labour supply model with preference formation

The main purpose of this section is to incorporate preference interdependence and habit formation into the neoclassical labour supply model described in Chapter 3. We allow for the fact that individuals may get used to working a certain number of hours per week (habit formation) and that the number of hours they prefer to work may depend on the number of hours other people worked (interdependence of preferences). Habit formation and preference interdependence will be referred to jointly as preference formation. We will derive the model for two-adult households, while the one-adult household is an evident simplification. We model preference formation by making the translation parameters δ_k for singles or δ_{mk} and δ_{fk} for households linearly dependent upon hours worked by other people in society, one period ago, and upon the number of hours worked by the individual himself (or herself) one period ago. Figure 4.1 gives a simplified example of the structure of this part of the model. We consider in this example a society with 5 individuals and we concentrate on the way the preferences of individual 3 are formed. The story starts in period 1. In that period we observe the numbers actually worked by all five individuals. Next we move to period 2 and see how the preferences of individual 3 change. In principle δ_3 in period 2 is affected by all the values of the actual numbers of hours in the previous period. However, not all of these exert the same amount of influence. The extent of the influence of each individual's behaviour (i.e. the number of hours worked) in the previous period is measured by the so-called reference weights $v_{k\ell}$

which indicate the weight attached by individual k to the behaviour of individual ℓ.

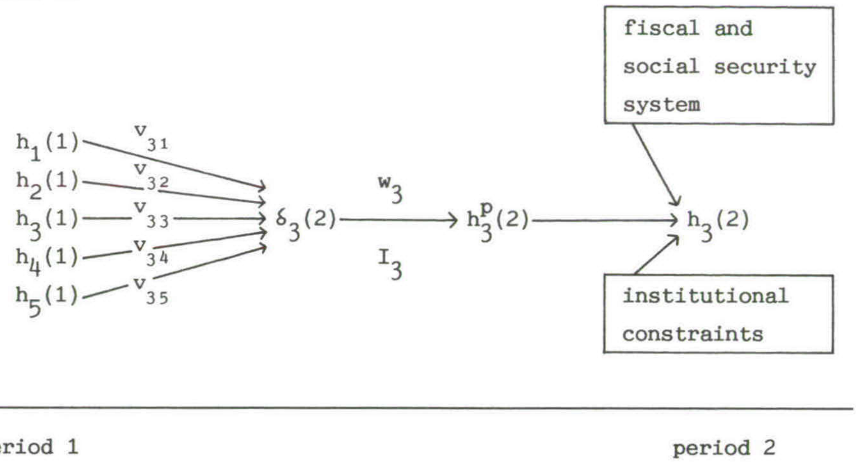


Figure 4.1 <u>Formation of the preferences of individual 3</u>

Thus v_{33} indicates the influence of habit formation (the effect of actual numbers of hours worked in the previous period on the current delta), v_{31} indicates the influence of the behaviour of individual 1 on δ_3; v_{32} gives the influence of individual 2, etc. Of course the deltas of individuals 1,2,4,5 are also influenced by what happened in the previous period, but in order not to mess up the graph, these influences are not shown in the figure.

Once δ_3 for this period is formed, the preferred number of hours h_3^p of individual 3 in period 2 follows from the budget constraint, i.e. from the wage rate w_k and from nonlabour income I_k. Generally, the preferred number of hours will not be realized. Rather the interplay of the individual's preferences, institutional constraints, and the fiscal and social security system generates a new value for actual hours h_3 in period 2.

Now in reality societies are quite a bit larger than the five-person society in the example of Figure 4.1. Thus in principle for each individual we might distinguish millions of reference weights to describe how his preferences are influenced by other people's behaviour. Obviously this is an impractical approach. Although we maintain the basic set-up sketched above, we impose more structure by making a number of additional assumptions. These assumptions are spelled out in detail below. First the basic structure of what the assumptions imply is sketched, Fig.4.2.

all males in society $\sum_{\ell\in P_m} v^{mm}_{k\ell}h_{m\ell}(-1)$	η_{mm}	male in family k $h_{mk}(-1)$	ζ_m	male in family k $h_{mk}(-1)$	ζ_m	male in family k $h_{mk}(-1)$	ζ_m
		other males $\hat{h}_{mmk}(-1)$	$1-\zeta_m$	own male friends $\sum_{\ell\in S_{mm}} q^{mm}_{k\ell}h_{m\ell}(-1)$	λ_{mm}	males in own social group $\bar{h}_{mmk}(-1)$	$1-\kappa_m$
						all males in society $\bar{\bar{h}}_m(-1)$	κ_m
				male friends of spouse $\sum_{\ell\in S_{fm}} q^{mm}_{k\ell}h_{m\ell}(-1)$	$1-\lambda_{mm}$	males in social group of spouse $\bar{h}_{fmk}(-1)$	$1-\kappa_m$
						all males in society $\bar{\bar{h}}_m(-1)$	κ_m
all females in society $\sum_{\ell\in P_f} v^{mf}_{k\ell}h_{f\ell}(-1)$	η_{mf}	females $\hat{h}_{mfk}(-1)$		own female friends $\sum_{\ell\in S_{mf}} q^{mf}_{k\ell}h_{f\ell}(-1)$	λ_{mf}	females in own social group $\bar{h}_{mfk}(-1)$	$1-\kappa_f$
						all females in society $\bar{\bar{h}}_f(-1)$	κ_f
				female friends of spouse $\sum_{\ell\in S_{ff}} q^{mf}_{k\ell}h_{f\ell}(-1)$	$1-\lambda_{mf}$	females in social group of spouse $\bar{h}_{ffk}(-1)$	$1-\kappa_f$
						all females in society $\bar{\bar{h}}_f(-1)$	κ_f

Figure 4.2 The partition of the reference group of the male in family k

Let us consider a two-adult household and more specifically the male in this household. Society is partitioned in a number of groups of different individuals and these all exert some influence on the male in the household under consideration. These groups are represented in Figure 4.2. First of all a distinction is made between the influence exerted by males and by females

$$\delta_{mk} = \delta_{m0} + \eta_{mm} \sum_{\ell \in P_m} v_{k\ell}^{mm} h_{m\ell}(-1) + \eta_{mf} \sum_{\ell \in P_f} v_{k\ell}^{mf} h_{f\ell}(-1) \qquad (4.1)$$

$$\delta_{fk} = \delta_{f0} + \eta_{ff} \sum_{\ell \in P_f} v_{k\ell}^{ff} h_{f\ell}(-1) + \eta_{fm} \sum_{\ell \in P_m} v_{k\ell}^{fm} h_{m\ell}(-1) \qquad (4.2)$$

where

P_m = the set of all adult males in society

P_f = the set all adult females in society

$h_{mk}(-1)$ = lagged value of actual hours worked by the male in household k

$h_{fk}(-1)$ = lagged value of actual hours worked by the female in household k

$\delta_{m0}, \delta_{f0}, \eta_{mm}, \eta_{mf}, \eta_{ff}, \eta_{fm}$ are parameters

$0 \leq \eta_{mm}, \eta_{mf}, \eta_{ff}, \eta_{fm} \leq 1$

$0 \leq v_{k\ell}^{mm}, v_{k\ell}^{mf}, v_{k\ell}^{ff}, f_{k\ell}^{fm} \leq 1$, for all k, ℓ

$\sum_{\ell \in P_m} v_{k\ell}^{mm}=1, \sum_{\ell \in P_f} v_{k\ell}^{mf}=1, \sum_{\ell \in P_f} v_{k\ell}^{ff}=1, \sum_{\ell \in P_m} v_{k\ell}^{fm}=1$, for all k

The $v_{k\ell}$'s are called <u>reference weights</u>. For example, $v_{k\ell}^{mf}$ represents the importance attached by the male in family k to labour supply behaviour of the female in family ℓ. If, for instance, $v_{k\ell}^{mf}=0$ the female in family ℓ does not belong to the reference group of the male in family k, if $v_{k\ell}^{mf}\neq 0$ she does. Similar interpretations can be given to $v_{k\ell}^{mm}$, $v_{k\ell}^{ff}$ and $v_{k\ell}^{fm}$. The parameters $\eta_{..}$ measure the importance of preference formation. For example, η_{mf} represents the extent to which the preference parameter δ_{mk} is influenced by the labour supply behaviour of females in the reference group of the male in family k. A necessary but not sufficient condition for the model is that the η's lie between zero and one.

By inserting equations (4.1) and (4.2) into (3.15) - (3.20) we obtain a model which relates observables to observables, but with far too many parameters to be estimated. To reduce the number of parameters a few assumptions are made.

First of all it is assumed that v_{kk}^{mm} and v_{kk}^{ff} do not vary with k, say $v_{kk}^{mm}=\zeta_m$ and $v_{kk}^{ff}=\zeta_f$. This means that the relative influence of habit formation is the same across households. The influence of the male's own past behaviour (habit formation), is depicted at the top of the Figure 4.2 by the parameter ζ_m.
Since $\sum_{\substack{\ell \in P_m \\ \ell \neq k}} v_{k\ell}^{mm}=1-\zeta_m$ and $\sum_{\substack{\ell \in P_f \\ \ell \neq k}} v_{k\ell}^{ff}=1-\zeta_f$, the relative influence of preference interdependence is also the same across households. The larger ζ_m (or ζ_f) the more important habit formation is relative to preference interdependence. Next we introduce new parameters $q_{k\ell}^{mm}$ defined by

$$\begin{aligned} q_{k\ell}^{mm} &= v_{k\ell}^{mm}/(1-\zeta_m) \qquad && k \neq \ell \\ &= 0 && k = \ell \end{aligned} \tag{4.3}$$

Obviously,

$$q_{k\ell}^{mm} \geq 0 \text{ for all } k,\ \ell,\ \sum_{\ell \in P_m} q_{k\ell}^{mm}=1 \text{ for all } k$$

Similarly, we define $q_{k\ell}^{ff}= v_{k\ell}^{ff}/(1-\zeta_f)$ for $k \neq \ell$ and zero otherwise. For notational simplicity we also replace $v_{k\ell}^{mf}$ and $v_{k\ell}^{fm}$ by $q_{k\ell}^{mf}$ and $q_{k\ell}^{fm}$ respectively.

Also the parameters $q_{k\ell}$ will be called "reference weights". We refer to the set of females for which $q_{k\ell}^{mf}>0$ as the female social reference group of the male in family k. Define

$$\hat{h}_{ijk} = \sum_{\ell \in P_j} q_{k\ell}^{ij} h_{j\ell} \qquad i,j = m,f \tag{4.4}$$

The quantities $\hat{h}_{ijk}$ $(i,j = m,f)$ are reference group means of working hours. For example, the quantity $\hat{h}_{mfk}$ is the mean of female hours in the reference group of the male in family k. This makes it possible to rewrite (4.1) and (4.2) as

$$\begin{aligned} \delta_{mk} = \delta_{m0} + \eta_{mm}[\zeta_m h_{mk}(-1) + (1-\zeta_m)\hat{h}_{mmk}(-1)] + \\ \eta_{mf}[\hat{h}_{mfk}(-1)] \end{aligned} \tag{4.5}$$

$$\begin{aligned} \delta_{fk} = \delta_{f0} + \eta_{ff}[\zeta_f h_{fk}(-1) + (1-\zeta_f)\hat{h}_{ffk}(-1)] + \\ \eta_{fm}[\hat{h}_{fmk}(-1)] \end{aligned} \tag{4.6}$$

with $0 \leq \zeta_m, \zeta_f \leq 1$

Next consider $\hat{h}_{ijk}$ in somewhat more detail. It seems reasonable to suppose that an individual will primarily assign positive reference weights to people whom he or she knows personally. Within the family we distinguish two channels through which one can get to know other people: The first channel is that one meets someone directly; the second channel is that one meets someone through his or her partner. Both these groups of people will probably influence the individuals' preferences, but possibly to a different degree. To formalize this idea, let us take $\hat{h}_{mmk}$ (that is the mean of male hours in the reference group of the husband in family k) and rewrite it as

$$\hat{h}_{mmk} = \sum_{\ell \in P_m} q_{k\ell}^{mm} h_{m\ell} = \sum_{\ell \in S_{mm}} q_{k\ell}^{mm} h_{m\ell} + \sum_{\ell \in S_{fm}} q_{k\ell}^{mm} h_{m\ell} \tag{4.7}$$

where S_{mm} is the set of males that the husband in family k has met directly and S_{fm} is the set of males he has met through his wife (males whom neither of the spouses have met are assumed to receive a reference weight equal to zero; thus they can be assigned arbitrarily to either of the two sets without loss of generality).

Let λ_{mm}^{k} be the total reference weight assigned to the males in S_{mm}, i.e.

$$\lambda_{mm}^{k} = \sum_{\ell \in S_{mm}} q_{k\ell}^{mm} \tag{4.8}$$

so that the total weight assigned to males in S_{fm} equals $1-\lambda_{mm}^{k}$. We assume that λ_{mm}^{k} is a drawing from a distribution with mean λ_{mm}, i.e.

$$\lambda_{mm}^{k} = \lambda_{mm} + \varepsilon_{mm}^{k} \tag{4.9}$$

where $E\left[\varepsilon_{mm}^{k}|\lambda_{mm}\right] = 0$. Next define $r_{k\ell}^{mm} = q_{k\ell}^{mm}/\lambda_{mm}$ and $s_{k\ell}^{mm} = q_{k\ell}^{mm}/(1-\lambda_{mm})$. As a result (4.7) can be rewritten as

$$\hat{h}_{mmk} = \lambda_{mm} \sum_{\ell \in S_{mm}} r_{k\ell}^{mm} h_{m\ell} + (1-\lambda_{mm}) \sum_{\ell \in S_{fm}} s_{k\ell}^{mm} h_{m\ell} + \text{error term} \tag{4.10}$$

Note that the $r_{k\ell}^{mm}$ and $s_{k\ell}^{mm}$ are non-negative and "on average" they add up to one, i.e. $E\left[\sum_{\ell \in S_{mm}} r_{k\ell}^{mm}|\lambda_{mm}\right] = E\left[\sum_{\ell \in S_{fm}} s_{k\ell}^{mm}|\lambda_{mm}\right] = 1$. We will also denote the $r_{k\ell}^{mm}$ and $s_{k\ell}^{mm}$ as "reference weights". Expressions analogous to (4.10) can be obtained for the other $\hat{h}_{ijk}$, i,j = m,f. The partitioning of the society in own male friends, male friends of the spouse, own female friends and female friends of the spouse with associated weights is shown in the third column of Figure 4.2. The weights λ_{mm} and $1-\lambda_{mm}$ give the relative weight of the two groups of males distinguished. Similarly, the weights λ_{mf} and $1-\lambda_{mf}$ only tell us something about the relative weights of the two groups of females distinguished. The parameters η_{mm} and η_{mf} tell us how the group of all males and the group of all females are weighed relative to one another. That is, how much are the preferences of the male in family k influenced by the behaviour of males and how much by the behaviour of females. Finally, the relative influences of habit formation and preference interdependence are denoted by the parameters ζ_m and $1-\zeta_m$ respectively.

Having thus distinguished the total influences of four groups of different individuals, there remains the task of saying something about the relative influence of the individuals within each of these groups. Here the crucial assumption made is that individuals primarily refer to similar others. To be more specific, we distinguish so-called social groups; these are groups of individuals who share certain characteristics, like education, and age. On the basis of some further assumptions it can be shown that the pattern of influence within each of the four groups distinguished can be approximated by a weighted sum of the influence of the people inside the social group and the influence of the people outside the social group. Finally, we have made assumptions on the distribution of reference weights within each of the elements of the partition. Here we closely follow Van de Stadt, Kapteyn, Van de Geer (1985), who make three assumptions which lead to the result that an expression like $\sum_{\ell \in S_{mm}} r_{k\ell}^{mm} h_{m\ell}$ can be approximated as

$$\sum_{\ell \in S_{mm}} r_{k\ell}^{mm} h_{m\ell} = (1-\kappa_m)\, \bar{h}_{mmk} + \kappa_m \bar{\bar{h}}_m + \text{error term} \tag{4.11}$$

where the error terms has mean zero and is independent of the other terms on the right hand side of (4.11); κ_m is a parameter, $0 \leq \kappa_m \leq 1$; $\bar{\bar{h}}_m$ is the mean number of hours worked by all males in society; $\bar{h}_{mmk}$ is the mean number of hours worked by all males in the social group of the male. The parameter κ_m is an indicator of how informative a social group is about the reference group of an individual. For instance, if $\kappa_m=0$ then equation (4.11) implies that the social group mean is a good indicator of the reference group mean. On the other hand if $\kappa_m=1$ the social group mean conveys no information whatsoever about the reference group mean.

Similar to (4.11), we obtain as an approximation for $\sum_{\ell \in S_{fm}} s_{k\ell}^{mm} h_{m\ell}$

$$\sum_{\ell \in S_{fm}} s_{k\ell}^{mm} h_{m\ell} = (1-\kappa_m)\, \bar{h}_{fmk} + \kappa_m \bar{\bar{h}}_m + \text{error term} \tag{4.12}$$

where $\bar{h}_{fmk}$ is the mean of male working hours in the social group of the female in family k (i.e. all males with characteristics equal to the characteristics of this female). Expressions similar to (4.11) and (4.12) are obtained for the various other reference group means that play a role in (4.3) and (4.4). This final partition is shown in the last column of Figure 4.2.

After the incorporation of the effect of household size according to (2.18) and (2.19), one obtains

$$\delta_{mk} = \delta_{m0} + \delta_{m1} f_k$$

$+ \eta_{mm}[\zeta_m(h_{mk}(-1) - \delta_{m1} f_{mk}(-1))$	"habit formation"
$+ (1-\zeta_m)\{\kappa_m(\bar{\bar{h}}_m(-1) - \delta_{m1}\bar{\bar{f}}(-1))$	"all males in society"
$+ \lambda_{mm}(1-\kappa_m)(\bar{h}_{mmk}(-1) - \delta_{m1}\bar{f}_{mk}((-1))$	"males in own social group"
$+ (1-\lambda_{mm})(1-\kappa_m)(\bar{h}_{fmk}(-1)-\delta_{m1}\bar{f}_{fk}(-1))\}]$	"males in social group of spouse"
$+ \eta_{mf}[\kappa_f(\bar{\bar{h}}_f(-1) - \delta_{m1}\bar{\bar{f}}(-1))$	"all females in society"
$+ \lambda_{mf}(1-\kappa_f)(\bar{h}_{mfk}(-1) - \delta_{m1}\bar{f}_{mk}(-1))$	"females in own social group"
$+ (1-\lambda_{mf})(1-\kappa_f)(\bar{h}_{ffk}(-1) -\delta_{m1}\bar{f}_{fk}(-1))]$	"females in social group of spouse"
$+ \psi_{mk}$	(4.13)

(Note that "own" refers to the male in family k)

$$\delta_{fk} = \delta_{f0} + \delta_{f1} f_k$$

$+ \eta_{ff}[\zeta_f(h_{fk}(-1) - \delta_{f1} f_k(-1))$	"habit formation"

$$+ (1-\zeta_f)\{\kappa_f(\bar{\bar{h}}_f(-1) - \delta_{f1}\bar{\bar{f}}(-1))$$ "all females in society"

$$+ \lambda_{ff}(1-\kappa_f)(\bar{h}_{ffk}(-1) - \delta_{f1}\bar{f}_{fk}(-1))$$ "females in own social group"

$$+ (1-\lambda_{ff})(1-\kappa_f)(\bar{h}_{mfk}(-1)-\delta_{f1}\bar{f}_{mk}(-1))\}]$$ "females in social group of spouse"

$$+ \eta_{fm}[\kappa_m(\bar{\bar{h}}_m(-1) - \delta_{f1}\bar{\bar{f}}(-1))$$ "all males in society"

$$+ \lambda_{fm}(1-\kappa_m)(\bar{h}_{fmk}(-1) - \delta_{f1}\bar{f}_{fk}(-1))$$ "males in own social group"

$$+ (1-\lambda_{fm})(1-\kappa_m)(\bar{h}_{mmk}(-1)-\delta_{f1}\bar{f}_{mk}(-1))]$$ "males in social group of spouse"

$$+ \psi_{fk} \qquad (4.14)$$

(Note that "own" refers to the female in family k)

where f_k = logarithm of the size of household k

$\bar{f}_{mk}$ = mean of the logarithm of the size of the households in the social group of the male in family k

$\bar{f}_{fk}$ = mean of the logarithm of the size of the households in the social group of the female in family k

$\bar{\bar{f}}$ = mean of the logarithm of the size of the households in society as a whole

ψ_{mk}, ψ_{fk} are random variables with mean zero and independent of the other variables on the right hand side of (4.13) - (4.14).

In Chapter 2 we saw already that the model without preference formation is overidentified. Thus it is sufficient to examine the delta-equations (4.13)-(4.14) for identification of the model with preference formation. In Appendix 4B we will show that the delta-equations are identified up to two parameters. If for example κ_m and κ_f are fixed, the remaining parameters are identified.

Although in the exposition we have concentrated on the explanation of the preferences of the male in a two-adult household, it should be clear that the modelling of preference formation for the female in the household is entirely analagous. The model for singles is a special case of the model for households. The only difference is that for a single there is only one channel through which he (or she) meets other people. We do not have to distinguish between people one meets directly and people one meets through one's spouse. For a single male the parameter δ is therefore specified as follows

$$\begin{aligned}
\delta_{mk} = {} & \delta_{m0} + \delta_{m1} f_k \\
& + \eta_{mm}[\zeta_m(h_{mk}(-1) - \delta_{m1} f_{mk}(-1)) && \text{"habit formation"} \\
& + (1-\zeta_m)\{\kappa_m(\bar{\bar{h}}_m(-1) - \delta_{m1}\bar{\bar{f}}(-1)) && \text{"all males in society"} \\
& + (1-\kappa_m)(\bar{h}_{mmk}(-1) - \delta_{m1}\bar{f}_{mk}(-1))\}] && \text{"males in own social group"} \\
& + \eta_{mf}[\kappa_f(\bar{\bar{h}}_f(-1) - \delta_{m1}\bar{\bar{f}}(-1)) && \text{"all females in society"} \\
& + (1-\kappa_f)(\bar{h}_{mfk}(-1) - \delta_{m1}\bar{f}_{mk}(-1)] && \text{"females in own social group"} \\
& + \psi_{mk} && (4.15)
\end{aligned}$$

For single females all "m's" become "f's" and the other way around. See Appendix 4B for a discussion of the identification of equation (4.15).

4.3 Estimation results

Equations (3.1)-(3.6) and (4.15) summarize the complete specification of the model for singles. The two-adult household model is specified by equations (3.15)-(3.20) and (4.13)-(4.14). Preferred hours is the endogenous variable. The likelihood functions are given in Chapter 3. For sample information about the main variables of interest, see Appendix 4A and Chapter 2. As mentioned in Section 4.2 the samples of individuals in the OSA and SEP survey have been partitioned in social groups in which the individuals have identical characteristics. The characteristics considered are age and education level (four and five categories, respectively). In Tables 4A.1 and 4A.2 (Appendix 4A) social group means are presented of working hours and log-family sizes lagged one year. Since we have no data on lagged family size in the OSA sample, we use current family size as a proxy. Generally, the numbers in the table conform to what one might expect; males work more hours per week than females, the older one gets the less one works, and the higher one's education the more hours one works (this holds especially for women). It is striking to see the sharp drop in mean hours when one moves to the older age categories and than especially for the lower educated. This is mainly caused by a corresponding drop of participation rates of these age categories.

The main results of the maximum likelihood estimation of the model for singles are presented in Table 4.1 and for two-adult households in Table 4.2.

A comparison of the log-likelihood values corresponding to the extended model and the standard model, introduced in Chapter 3 makes it clear that preference formation is a highly significant factor in labour supply. Yet, the estimation results for the extended model are not uniformly satisfactory. In almost all cases the estimates for ζ_m and ζ_f are close to one or had to be constrained by an upper limit (one), except for single females in the OSA-data set. For both male and female the importance of habit formation relative to preference interdependence is overwhelming. Also the estimates of η_{ff} had to be constrained by the upper limit one in most cases, which implies that preference formation is an important variable in female labour supply. Females in families do not refer to males (η_{fm}=0). Male labour supply appears to be less influenced by preference formation than female labour supply (the sum of η_{mm} and η_{mf} is much smaller than the sum of η_{ff} and η_{fm} for all groups). Furthermore, both men and women refer more to women (η_{mf}, η_{ff}) than to men (η_{mm}, η_{fm}). A value of λ_{mf} of about 0.5 implies that men refer to women of their own age and education as well as to women of the same age and education as their wives. A λ_{ff} and λ_{mm} of 1 mean that a (fe)male's social group only consists of (wo)men of her/his own age and education, and not of (wo)men of the same age and education as the partner. Generally, social group means of male hours seem to have little impact. One explanation for this may be that male hours in The Netherlands are pretty much institutionally determined. The variation in social group means of male hours, shown in Table 4A.1, therefore mainly reflects variation across the life-cycle, across education levels of school enrolment, involuntary unemployment, and disability (for instance, in the age group 45-65, 37 percent of the male population in The Netherlands is on disability). The households whose labour supply we are trying to explain contain one or two able-bodied spouses who are available for market work. These households' reference groups may hardly contain the school-going, students, the disabled persons or even the unemployed. If this is true, the use of social group means to approximate reference group means may be a poor practice. Without data that contains more specific information about reference groups there is not too much we can do. For females the situation is less bleak, because their working hours are less affected by institutional constraints and hence the observed variation in social group means are probably more representative of variation in reference group means.

Table 4.4 summarizes the "total" influence, i.e. the reduced form coefficients, of habit formation and preference interdependence on the parameters δ_{mk} and δ_{fk} (compare equations (4.5) and (4.6)). The numbers are invariant to the arbitrary choice of the unidentified parameters κ_m and κ_f which we fixed at 0.5 to identify the remaining parameters. The small influence of preference interdependence is clear from this table. Yet, as will be seen below, preference interdependence does make a significant contribution. Only for single females in the OSA-data set preference interdependence has a larger influence than habit formation.

Turning now to the interpretation of the estimated values of the other parameters, we note that both columns of Table 4.1 and 4.2 show

negative income effects, implying leisure to be a normal good, and positive own linear wage effects. It is striking to see that for all different groups both the income effects and the own wage effects are smaller in absolute value than in the standard model (compare Tables 3.8 and 3.10 in Chapter 3). This could be explained by the fact that part of the influence of preference formation is taken up by the wage and income variables in the standard model. This would mean that in that case the wage and income effects are overestimated. The smaller wage effects are confirmed by Table 4.3 (compare Table 3.11). Moreover, we see that for single females in the OSA-data family size has a negative effect on the labour supply. The reduced form coefficient of log family size is -6. The value of δ_{f1} (-623) in the OSA two-adult household model is due to the fact that ζ_f is close to 1 and η_{ff} is equal to one. In fact the reduced form coefficient for (log) family size is
$\delta_{f1}(1-\eta_{ff}\zeta_f) + \beta_f\left[\delta_{m1}\left[1-\eta_{mm}\zeta_m\right]w_m + \delta_{f1}\left[1-\eta_{ff}\zeta_f\right]w_f\right] = -12$ (See Appendix 4B). This means that if a family of two is extended to three persons, the wife works 4.9 hours less per week. A second child reduces the wife's working hours by an additional 3.5 hours per week. In the standard model the first child reduces the number of working hours of the wife by 19, the second child by 14 additional hours. Part of the explanation of the large difference between the two models could be the fact that the extended model fails to capture the aspect that a woman who gets her first child will get to know more women with children, so that the group average of hours worked will be reduced. On the other hand, in the standard model there is no reference at all. Both for families and single females in the SEP-data δ_{f1} cannot be identified because $\zeta_f=1$ and $\eta_{mf}=0$.

In Figures 4.3 and 4.4 the short term labour supply curves are drawn. Both for the OSA- and the SEP-data the labour supply curves are flatter than those in the standard model (see Figures 3.4 and 3.5). In this so-called extended model we have to distinguish between short term and long term supply curves. The short term supply curve describes the relation between one's wage rate and the number of hours a person wants to work, assuming that nothing else (including the partner's actual number of hours) changes. In the long run supply curve we account for habit formation and for the fact that behaviour of one individual affects behaviour of others. The conditions under which long run stability of the dynamic model is guaranteed are investigated in Appendix 4C. We only present short term supply curves, for long term curves is referred to Kapteyn, Woittiez and ten Hacken (1989).

In Figures 4.5 and 4.6 the simulated preferred hours distributions are shown. The simulated hours distributions show a relatively high percentage of females in families working only a few hours and a low percentage of nonparticipants. As a result the labour supply curves of females in families lie a bit higher than in the standard model. For single females in the SEP the simulated hours distribution shows besides a peak at 0 also a small and thick peak around 36 hours, in contrast with the standard model. For the rest the patterns of the figures are the same as in Chapter 3.

In Tables 4.5-4.10 more estimation results are presented for two different specifications of the model. Tables 4.5-4.7 show that the hypothesis $\zeta_m=1$ and/or $\zeta_f=1$ (no preference interdependence) can be rejected at the 0.5% level for single females but not for married females in the OSA-data. For all other groups the parameters ζ_m and ζ_f reached their upper bound of one when left free. Consequently, the hypothesis of no preference interdependence ($\zeta_m=\zeta_f=1$) cannot be rejected. The hypothesis of no habit formation ($\zeta_m=\zeta_f=0$) is rejected decisively (Tables 4.8-4.10). Comparing the elasticities of the various versions, one can see that the wage elasticities in the model with only habit formation are smaller than in the model with only preference interdependence.

Table 4.1 Estimation results for the extended model, one-adult households[a)]

OSA

	males			females	
δ_0	11	(14)		-41	(17)
β	-0.01	(0.01)		-0.02	(0.004)
δ_1	5	(5)		-10	(5)
γ	0.1	(0.4)		1.0	(0.2)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	1	(u.b.)	ζ_f	0.29	(0.05)
η_{mm}	0.23	(0.05)	η_{ff}	1	(u.b.)
η_{mf}	0	(ℓ.b.)	η_{fm}	0.83	(0.39)
σ_h	7.6	(0.5)		11.9	(0.8)
σ_v	4437	(24816)		450	(307)
log lik.	-388.9			-538.1	

SEP

	males			females	
δ_0	1	(9)		-12	(3)
β	-0.007	(0.004)		-0.014	(0.003)
δ_1	-2	(8)		-	-
γ	0.1	(0.5)		0.7	(0.3)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	1	(u.b.)	ζ_f	1	(u.b.)
η_{mm}	0.57	(0.06)	η_{ff}	1	(u.b.)
η_{mf}	0.3	(0.3)	η_{fm}	0	(ℓ.b.)
σ_h	13.3	(0.8)		10.2	(0.4)
σ_v	1.10^5	(1.10^6)		411	(194)
log lik.	-493.0			-345.4	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses
It is not possible to identify the parameters κ_m, κ_f. We assume both to be equal to 0.5.

Table 4.2 Estimation results for the extended model, two-adult households[a)]

	OSA		SEP	
α	-0.0002	(0.003)	-0.04	(0.04)
β_m	-0.0004	(0.0004)	-0.009	(0.001)
β_f	-0.0002	(0.0001)	-0.000001	(u.b.)
γ_m	0.01	(ℓ.b.)	1.4	(1.5)
γ_f	0.8	(0.4)	1.1	(0.3)
δ_{m1}	1.9	(0.9)	-22	(4)
δ_{f1}	-623	(2402)	-	-
δ_{m0}	30	(20)	71	(148)
δ_{f0}	-18	(9)	-21	(3)
ϑ	7000	(4.10^5)	11373	(17202)
ζ_m	1	(u.b.)	1	(u.b.)
ζ_f	0.98	(0.05)	1	(u.b.)
η_{mm}	0.20	(0.01)	0.81	(0.02)
η_{mf}	0.22	(0.09)	1	(u.b.)
η_{ff}	1	(u.b.)	1	(u.b.)
η_{fm}	0	(ℓ.b.)	0	(ℓ.b.)
λ_{mm}	1	(u.b.)	1	(u.b.)
λ_{mf}	0.58	(0.30)	0.38	(0.18)
λ_{ff}	1	(u.b.)	1	(u.b.)
λ_{fm}	-	-	-	-
σ_m	6.0	(0.1)	8.5	(0.1)
σ_f	12.7	(0.5)	12.4	(0.3)
ρ	-0.08	(0.04)	0.03	(0.06)
σ_{mv}	3.10^5	(1.10^6)	2044	(2141)
σ_{fv}	2.10^6	(3.10^7)	3803	(40000)
log lik.	-4114.5		-4735.0	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses
It is not possible to identify the parameters κ_m, κ_f. We assume both to be equal to 0.5.

Table 4.3 Wage elasticities for the extended model[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	-0.03	-0.00	-0.01	0.02
	(0.13)	(0.01)	(0.22)	(0.04)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.11	0.39	0.11	0.69
	(0.11)	(0.51)	(0.24)	(0.63)

a) Standard errors in parentheses

Table 4.4: Total influence of habit formation and preference interdependence[a)]

	OSA			
	singles		families	
	influence on			
influence of	δ_{mk}	δ_{fk}	δ_{mk}	δ_{fk}
habit formation, $h_{mk}(-1)$	0.23	-	0.20	-
habit formation, $h_{fk}(-1)$	-	0.29	-	0.98
mean of h_m in r.g. of male, $\hat{h}_{mmk}(-1)$	0	-	0	-
mean of h_f in r.g. of male, $\hat{h}_{mfk}(-1)$	0	-	0.22	-
mean of h_f in r.g. of female, $\hat{h}_{ffk}(-1)$	-	0.71	-	0.02
mean of h_m in r.g. of female, $\hat{h}_{fmk}(-1)$	-	0.83	-	0

	SEP			
	singles		families	
	influence on			
influence of	δ_{mk}	δ_{fk}	δ_{mk}	δ_{fk}
habit formation, $h_{mk}(-1)$	0.57	-	0.81	-
habit formation, $h_{fk}(-1)$	-	1	-	1
mean of h_m in r.g. of male, $\hat{h}_{mmk}(-1)$	0	-	0	-
mean of h_f in r.g. of male, $\hat{h}_{mfk}(-1)$	0.3	-	1	-
mean of h_f in r.g. of female, $\hat{h}_{ffk}(-1)$	-	0	-	0
mean of h_m in r.g. of female, $\hat{h}_{fmk}(-1)$	-	0	-	0

a) r.g.= reference group

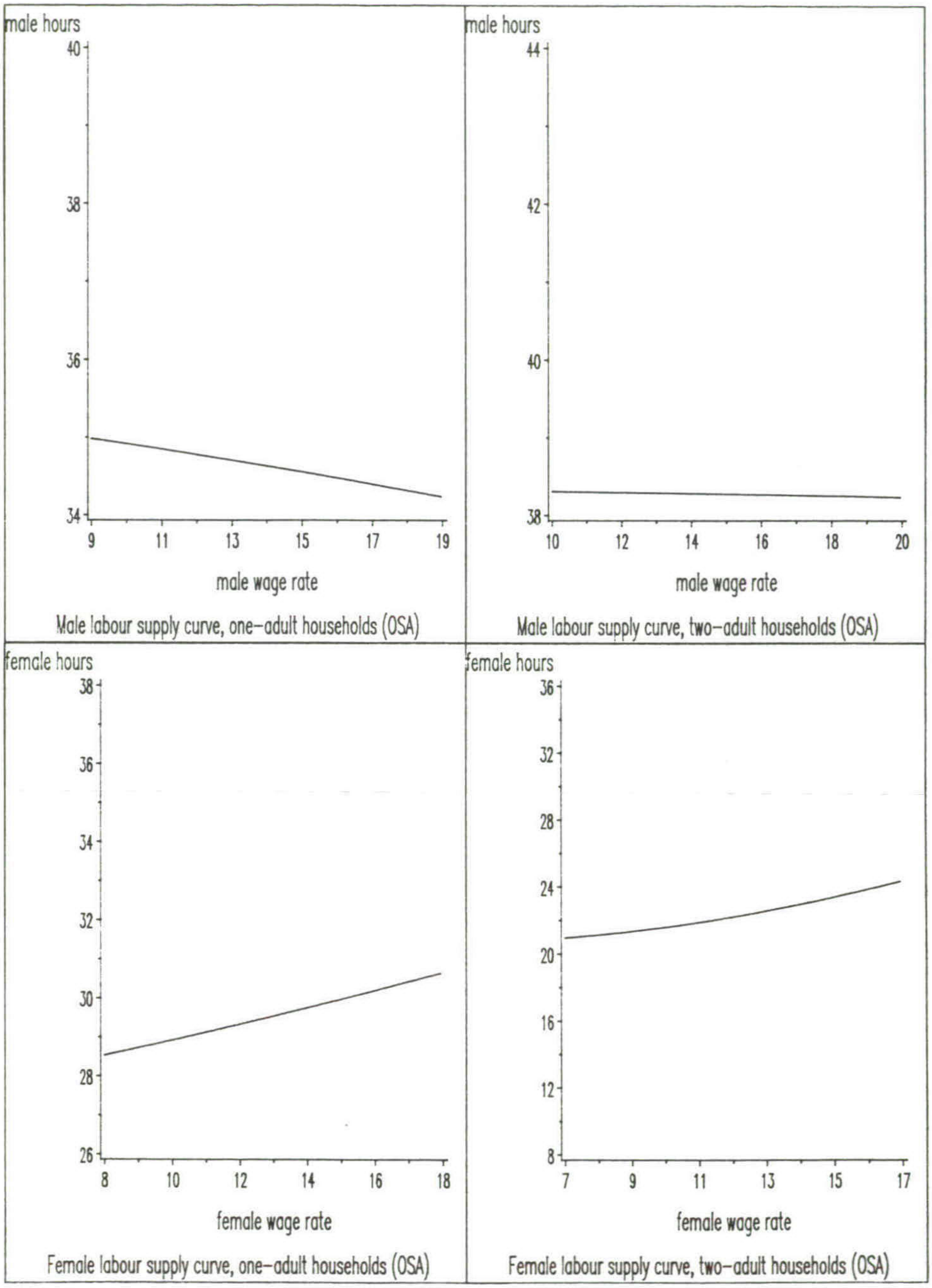


Figure 4.3 Labour supply curves (OSA)

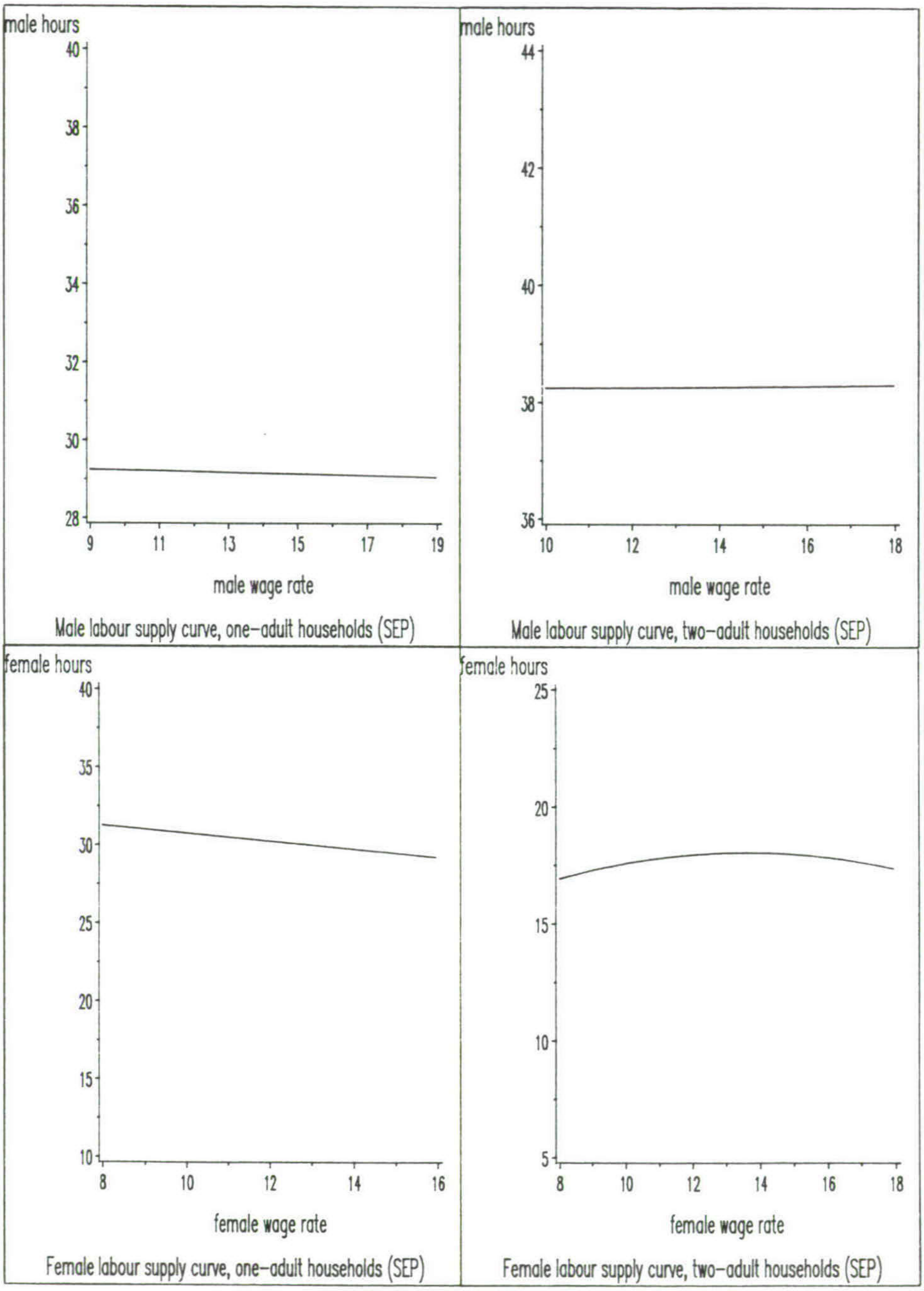


Figure 4.4 Labour supply curves (SEP)

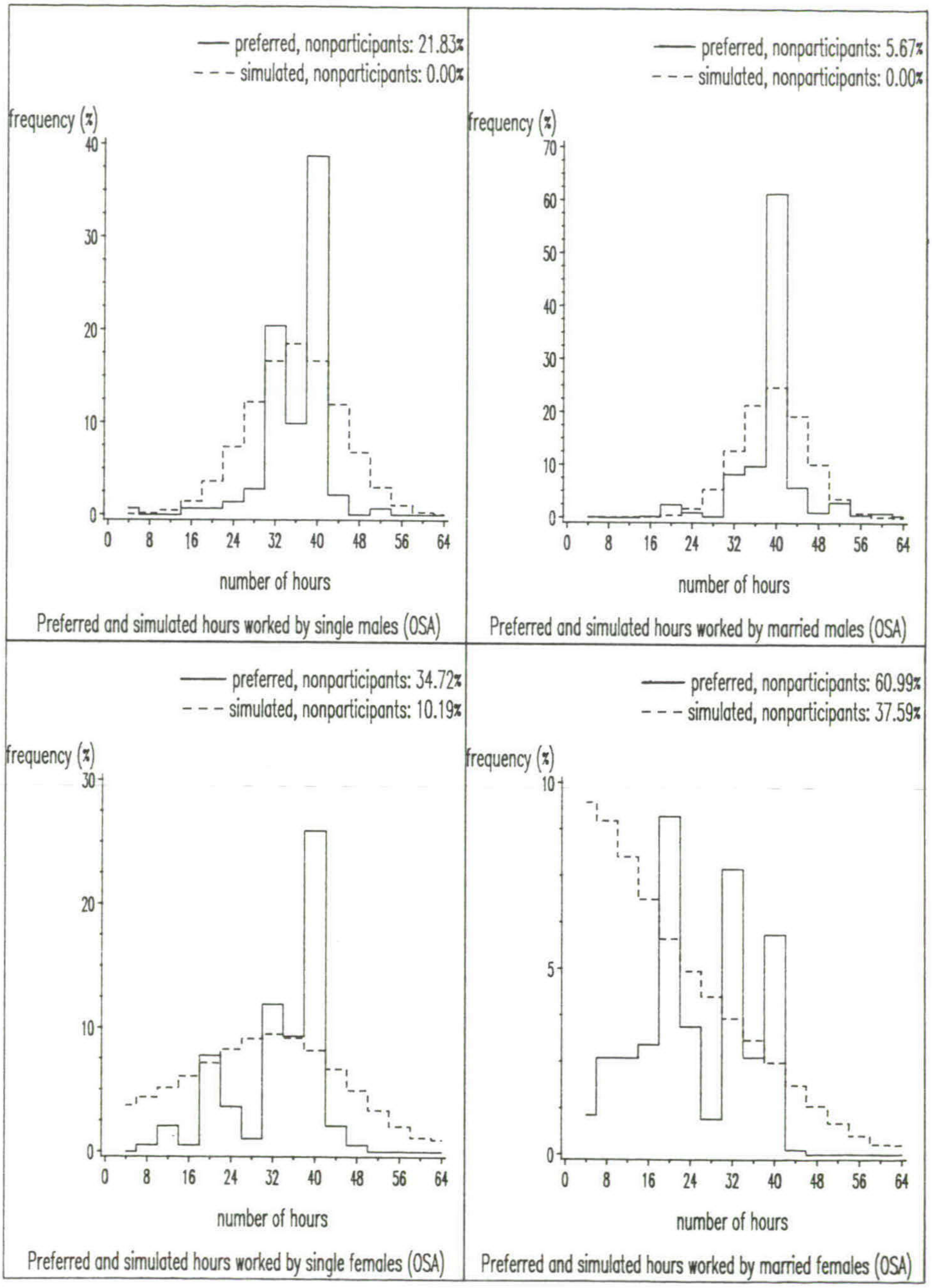


Figure 4.5 Hours distributions (OSA)

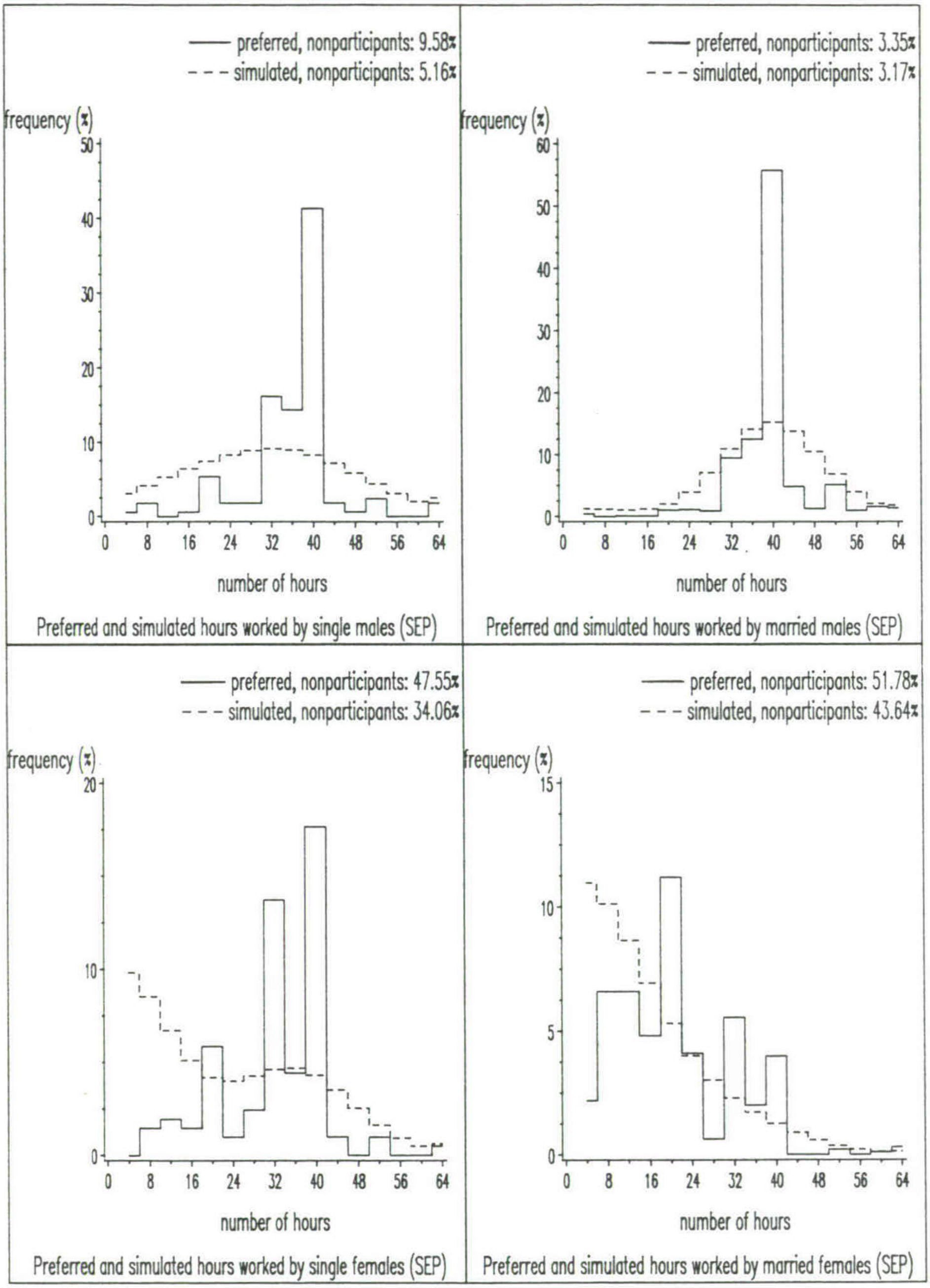


Figure 4.6 <u>Hours distributions (SEP)</u>

Table 4.5 Estimation results for the extended model, only habit formation, one-adult households[a)]

OSA

	males			females	
δ_0	11	(14)		-11	(15)
β	-0.01	(0.01)		-0.02	(0.004)
δ_1	5	(5)		-9	(4)
γ	0.1	(0.4)		0.7	(0.1)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	1	(fixed)	ζ_f	1	(fixed)
η_{mm}	0.23	(0.05)	η_{ff}	0.38	(0.06)
η_{mf}	0	(ℓ.b.)	η_{fm}	0.55	(0.31)
σ_h	7.6	(0.5)		13.0	(0.8)
σ_v	4437	(24816)		90	(43)
log lik.	-388.9			-541.4	

SEP

	males			females	
δ_0	1	(9)		-12	(3)
β	-0.007	(0.004)		-0.014	(0.003)
δ_1	-2	(8)		-	-
γ	0.1	(0.5)		0.7	(0.3)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	1	(fixed)	ζ_f	1	(fixed)
η_{mm}	0.57	(0.06)	η_{ff}	1	(u.b.)
η_{mf}	0.3	(0.3)	η_{fm}	0	(ℓ.b.)
σ_h	13.3	(0.8)		10.2	(0.4)
σ_v	1.10^5	(1.10^6)		411	(194)
log lik.	-493.0			-345.4	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses.
It is not possible to identify the parameters $\varkappa_m$, $\varkappa_f$. We assume both to be equal to 0.5.

Table 4.6 Estimation results for the extended model, only habit formation, two-adult households[a)]

	OSA		SEP	
α	-0.0002	(0.003)	-0.04	(0.04)
β_m	-0.0004	(0.0004)	-0.009	(0.001)
β_f	-0.0002	(0.0001)	-0.000001	(u.b.)
γ_m	0.01	(ℓ.b.)	1.4	(1.5)
γ_f	0.6	(0.4)	1.1	(0.3)
δ_{m1}	1.9	(0.9)	-22	(4)
δ_{f1}	-980	(6821)	-	-
δ_{m0}	30	(20)	71	(148)
δ_{f0}	-17	(10)	-21	(3)
ϑ	6000	(4.10^5)	11373	(17202)
ζ_m	1	(fixed)	1	(fixed)
ζ_f	1	(fixed)	1	(fixed)
η_{mm}	0.20	(0.01	0.81	(0.02)
η_{mf}	0.22	(0.09)	1	(u.b.)
η_{ff}	0.99	(0.05)	1	(u.b.)
η_{fm}	0.01	(0.05)	0	(ℓ.b.)
λ_{mm}	-	-	1	(u.b.)
λ_{mf}	0.58	(0.31)	0.38	(0.18)
λ_{ff}	-	-	1	(u.b.)
λ_{fm}	1	(u.b.)	-	-
σ_m	6.0	(0.1)	8.5	(0.1)
σ_f	12.7	(0.5)	12.4	(0.3)
ρ	-0.08	(0.04)	0.03	(0.06)
σ_{mv}	3.10^6	(1.10^7)	2044	(2141)
σ_{fv}	1.10^7	(2.10^8)	3803	(40000)
log lik.	-4115.3		-4735.0	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses
It is not possible to identify the parameters κ_m, κ_f. We assume both to be equal to 0.5.

Table 4.7 <u>Wage elasticities for the extended model, only habit formation</u>[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m} \frac{w_m}{h_m}$	-0.03	-0.00	-0.01	0.02
	(0.13)	(0.01)	(0.22)	(0.04)
$\frac{\partial h_f}{\partial w_f} \frac{w_f}{h_f}$	0.05	0.33	0.11	0.69
	(0.06)	(0.53)	(0.24)	(0.63)

a) Standard errors in parentheses

Table 4.8 Estimation results for the extended model, only preference interdependence, one-adult households[a)]

OSA

	males			females	
δ_0	17	(18)		-47	(12)
β	-0.01	(0.01)		-0.03	(0.004)
δ_1	4	(4)		-10	(4)
γ	0.5	(0.3)		1.5	(0.2)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	0	(fixed)	ζ_f	0	(fixed)
η_{mm}	0	(ℓ.b.)	η_{ff}	0.73	(0.45)
η_{mf}	0	(ℓ.b.)	η_{fm}	1	(u.b.)
σ_h	8.3	(0.6)		13.0	(0.9)
σ_v	971	(1250)		236	(116)
log lik.	-400.3			-546.9	

SEP

	males			females	
δ_0	19	(10)		-66	(18)
β	-0.09	(0.003)		-0.05	(0.01)
δ_1	4	(3)		-6	(9)
γ	0.3	(0.6)		2.3	(0.4)
ϑ	-1650	(fixed)		-355	(fixed)
ζ_m	0	(fixed)	ζ_f	0	(fixed)
η_{mm}	0	(ℓ.b.)	η_{ff}	1	(u.b.)
η_{mf}	0	(ℓ.b.)	η_{fm}	1	(u.b.)
σ_h	16.3	(0.9)		19.9	(1.6)
σ_v	5.10^5	(1.10^6)		154	(35)
log lik.	-518.8			-404.6	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses
It is not possible to identify the parameters κ_m, κ_f. We assume both to be equal to 0.5

Table 4.9 Estimation results for the extended model, only preference interdependence, two-adult households[a)]

	OSA		SEP	
α	0.7	(0.6)	1.2	(0.2)
β_m	-0.001	(0.004)	-0.020	(0.002)
β_f	-0.01	(0.005)	-0.009	(0.004)
γ_m	0.15	(0.19)	0.9	(0.2)
γ_f	8.6	(6.6)	3.6	(0.6)
δ_{m1}	1.5	(0.9)	-24	(4)
δ_{f1}	-45	(5)	-50	(5)
δ_{m0}	103	(53)	-12	(4)
δ_{f0}	682	(730)	-24	(11)
ϑ	7.10^5	(1.10^6)	435	(216)
ζ_m	0	(fixed)	0	(fixed)
ζ_f	0	(fixed)	0	(fixed)
η_{mm}	0	(ℓ.b.)	0	(ℓ.b.)
η_{mf}	0.19	(0.01)	1	(u.b.)
η_{ff}	1	(u.b.)	0	(ℓ.b.)
η_{fm}	0	(ℓ.b.)	0	(ℓ.b.)
λ_{mm}	-	-	-	-
λ_{mf}	0.42	(0.41)	0.60	(0.19)
λ_{ff}	-	-	-	-
λ_{fm}	0.14	(0.30)	-	-
σ_m	6.3	(0.1)	12.3	(0.2)
σ_f	20.8	(1.3)	26.3	(1.1)
ρ	-0.12	(0.05)	-0.73	(0.02)
σ_{mv}	6000	(9800)	3311	(815)
σ_{fv}	3209	(29646)	20464	(63370)
log likelihood	-4380.6		-5196.4	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses
It is not possible to identify the parameters κ_m, κ_f. We assume both to be equal to 0.5.

Table 4.10 Wage elasticities for the extended model, only preference interdependence[a)]

	OSA		SEP	
	one-adult	two-adult	one-adult	two-adult households
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.09	0.01	0.03	-0.03
	(0.12)	(0.03)	(0.20)	(0.06)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.22	0.53	0.65	1.86
	(0.09)	(0.64)	(0.42)	(3.09)

a) Standard errors in parentheses

4.4 A factor analytic model of reference groups

In Section 2 of this chapter social reference group variables were constructed on the basis of a set of assumptions. Given the assumptions, the significance of reference group influences on the labour supply of households is corroborated by the data. Since the SEP-data contain answers to a number of direct questions about reference groups of individuals, it is possible to have a more sophisticated look at the influence of reference groups. In this section the direct information on reference groups is used to construct a so-called latent variables model of the influence of reference groups on labour supply.

It is the goal of this section to employ the direct information on reference groups within the formal structure of the labour supply model developed in the preceding chapters. The basic idea is to use the answers to the questions about reference groups to construct a number of indicators of the family size and number of hours worked in the reference group of each respondent. These indicators are next used in a so-called confirmatory factor analysis model to estimate the relation of the indicators to the "true" family size and number of hours worked in the reference group. These true variables can then be proxied on the basis of the observed indicators. The proxies are used in the household labour supply model instead of the original variables, which were constructed on the basis of education and age. The factor analytic model will be explained in this section and the results of its estimation will be given in Section 4.5. In Section 4.5 we also present the results of reestimation of the labour supply model with the new reference group variables.

For clarity of exposition we will first present the questions that were asked in the survey about one's reference groups. The questionnaire administered to the SEP-respondents included the following questions

1. "The following questions are about your social environment, that is people whom you meet frequently, like friends, neighbours, acquaintances or possibly people you meet at work. Thinking of the people in your social environment, can you indicate to which age class they belong primarily? Choose the answer which is most in accordance with reality."

[01] under 16
[02] 16-20
[03] 21-25
[04] 26-30
[05] 31-35
[06] 36-40
[07] 41-45
[08] 46-50
[09] 51-55
[10] 56-60
[11] 61-65
[12] 65-70
[13] 71 and over

2 "People in your social environment may live alone or live in a multi-person household (e.g. with partner and children). What is the typical size of the number of persons in households in your social environment? Check number one if most households in your social environment consist of single persons, check number two if most households in your social environment have two persons, etc."

[1] one person
[2] two persons
[3] three persons
[4] four persons
[5] five persons
[6] six persons or more

3 "What educational level do most people in your social environment have? Check the number corresponding with the answer that comes closest to reality."

[1] primary education
[2] lower vocational education
[3] intermediate general education
[4] high school
[5] intermediate vocational education
[6] higher vocational education B.A. or B.SC.

[7] masters degree

The purpose of these questions was to collect some direct information on the composition of the reference group of each respondent. The decision to include these questions in the questionnaire was made in the context of a research project on subjective poverty, in which it was hypothesized that the incomes and family composition of one's reference group are important determinants of one's subjective well-being. For the purpose of explaining household labour supply on the basis of preference interdependence the questions suffer from one serious defect, because no questions have been asked about the number of hours worked in the reference group. Although this certainly detracts from the value of the information, it still is a lot more direct than what we have used in the previous chapters. Let us now describe the construction of the indicators and then explain the way the indicators are related to the "true" reference group variables. For singles we construct four indicators for hours worked in a respondent's reference group and five indicators for the family size. The first two indicators for reference group hours are simply the ones used in the previous chapters, i.e. the mean number of hours worked by all males or females in the sample who share the characteristics educational level and age bracket with the respondent. The construction of the third and fourth indicator is analogous, but now it is the mean number of hours of all males or females in the sample who have the educational level and age indicated by the respondent as typical for his or her reference group.

For family size the first four indicators are constructed analogously to the variables for hours, namely as sample means for people with certain characteristics. The fifth indicator for family size is the answer to question 2 above. For two-adult households we have four more indicators for hours worked as well as for family size. Besides the abovementioned indicators we use also the mean family size and the mean number of hours worked by all males or females in the sample who share the characteristics education level and age bracket with the partner of the respondent. And we use the mean number of hours of all males or females in the sample who have the education level and age indicated by the partner of the respondent as typical for his or her reference group.

Having constructed the indicators, how can we use them to find reasonable operationalizations of the "true" reference group variables? First we restrict our attention to singles. Let x_1 through x_4 be the indicators of hours worked in the reference group of some respondent, and let ξ_{hm} and ξ_{hf} be the true value of the number of hours worked by males and females respectively in the reference group. Then we assume the following relationship between these variables

$$x_1 = \alpha_1 \xi_{hm} + \nu_1 \tag{4.16}$$

$$x_2 = \alpha_2 \xi_{hm} + \nu_2 \tag{4.17}$$

$$x_3 = \alpha_3 \xi_{hf} + \nu_3 \tag{4.18}$$

$$x_4 = \alpha_4 \xi_{hf} + \nu_4 \tag{4.19}$$

where ν_1 through ν_4 are random terms, normally distributed with mean zero and mutually independent

x_1 = mean number of hours worked by all males who are in the same age and education bracket as the respondent

x_2 = mean number of hours worked by all males who have the age and education level indicated by the respondent as typical for his or her reference group

x_3 = mean number of hours worked by all females who are in the same age and education bracket as the respondent

x_4 = mean number of hours worked by all females who have the age and education level indicated by the respondent as typical for his or her reference group

Thus one sees that the indicators are assumed to be dependent on the two true variables, but their coefficients may differ and the relations suffer from errors, ν_1 through ν_4. Technically, (4.16)-(4.19) is a factor analysis model, where ξ_{hm} and ξ_{hf} are the factors, α_1 through α_4 are the factor loadings, x_1 through x_4 are the specific indicators.
For the five indicators of family size a similar model is assumed

$$x_5 = \alpha_5 \xi_{fs} + \nu_5 \qquad (4.20)$$

$$x_6 = \alpha_6 \xi_{fs} + \nu_6 \qquad (4.21)$$

$$x_7 = \alpha_7 \xi_{fs} + \nu_7 \qquad (4.22)$$

$$x_8 = \alpha_8 \xi_{fs} + \nu_8 \qquad (4.23)$$

$$x_9 = \alpha_9 \xi_{fs} + \nu_9 \qquad (4.24)$$

where x_5 through x_8 are analogues of x_1 through x_4 with hours worked replaced by family size and the error terms ν_5 to ν_9 are random terms, normally distributed and mutually independent.
x_9 = answer to question 2 above

For estimation of the two models we need the values of all nine indicators, introduced in equations (4.16)-(4.24). This leaves us with 54 observations for single males and 153 for single females. We have estimated the two different models seperately for males and for females.

Obviously, estimation of the two models for two-adult households with all the indicators included requires that all the questions have been answered by both spouses in the households. It turns out that the routing in the questionnaire of the SEP has been such that the questions have only been asked to individuals who have their own independent source of income. As a result, many observations had to be discarded, including most households with nonworking housewives. For estimation only 318 households remain.

For families, we exploit the information provided by husbands and wives simultaneously. We do this by estimation of two models. The first model consists of four equations analagous to equations (4.16)-(4.19) for both husband and wife, so that we have eight indicators and two "true" reference group variables

$$x_1^m = \alpha_1^m \xi_{hm} + \nu_1 \qquad (4.25)$$

$$x_2^m = \alpha_2^m \xi_{hm} + \nu_2 \tag{4.26}$$

$$x_1^f = \alpha_1^f \xi_{hm} + \nu_3 \tag{4.27}$$

$$x_2^f = \alpha_2^f \xi_{hm} + \nu_4 \tag{4.28}$$

$$x_3^m = \alpha_3^m \xi_{hf} + \nu_5 \tag{4.29}$$

$$x_4^m = \alpha_4^m \xi_{hf} + \nu_6 \tag{4.30}$$

$$x_3^f = \alpha_3^f \xi_{hf} + \nu_7 \tag{4.31}$$

$$x_4^f = \alpha_4^f \xi_{hf} + \nu_8 \tag{4.32}$$

For the definition of the variables x^m, we must substitute in the definition of x "respondent" by "male in the household". Likewise x^f is defined for the female in the household. The second model consists of analogues of (4.20)-(4.24) for both husband and wife jointly. Since in the labour supply model there is no distinction between reference group family size for male and female, it is a natural approach to take the ten indicators as functions of the same underlying "true" reference group mean

$$x_5^m = \alpha_5^m \xi_{fs} + \nu_9 \tag{4.33}$$

$$x_6^m = \alpha_6^m \xi_{fs} + \nu_{10} \tag{4.34}$$

$$x_7^m = \alpha_7^m \xi_{fs} + \nu_{11} \tag{4.35}$$

$$x_8^m = \alpha_8^m \xi_{fs} + \nu_{12} \tag{4.36}$$

$$x_9^m = \alpha_9^m \xi_{fs} + \nu_{13} \tag{4.37}$$

$$x_5^f = \alpha_5^f \xi_{fs} + \nu_{14} \tag{4.38}$$

$$x_6^f = \alpha_6^f \xi_{fs} + \nu_{15} \tag{4.39}$$

$$x_7^f = \alpha_7^f \xi_{fs} + \nu_{16} \tag{4.40}$$

$$x_8^f = \alpha_8^f \xi_{fs} + \nu_{17} \tag{4.41}$$

$$x_9^f = \alpha_9^f \xi_{fs} + \nu_{18} \tag{4.42}$$

In Appendix 4D the correlation matrices of the indicators are given.

Table 4.11 presents the results of the factor analyses performed for family size and hours respectively. The estimation method employed is maximum likelihood, using the well-known LISREL program.

Table 4.11. Estimation results for the confirmatory factor analyses

	singles				families			
	males		females		males		females	
Par.	est.	stand. error	est.	stand. error	est.	stand. error	est.	stand. error
				HOURS				
α_1	4.38	1.25	2.75	0.52	1.69	0.21	1.87	0.25
α_2	3.84	0.60	5.10	0.63	3.57	0.44	1.82	0.27
α_3	1.62	0.61	5.08	0.77	2.35	0.20	4.35	0.22
α_4	4.74	0.79	5.12	0.53	3.23	0.21	4.51	0.22
R^2	0.99		0.93			0.97		
				FAMILY SIZE				
α_5	0.15	0.02	0.20	0.01	0.11	0.01	0.11	0.01
α_6	0.06	0.03	0.09	0.02	0.10	0.01	0.14	0.01
α_7	0.19	0.03	0.17	0.01	0.06	0.01	0.10	0.01
α_8	0.07	0.04	0.08	0.01	0.08	0.01	0.11	0.01
α_9	0.15	0.02	0.08	0.03	0.15	0.02	0.11	0.02
R^2	0.75		0.88			0.73		
corr. between male and female ref. group hours						0.52		

The estimation results are satisfactory from a statistical viewpoint. The R^2's, indicating the extent of the multiple correlation between the "true" reference group variables and the indicators range from 0.52 to 0.99. The signs of the α's are all positive, as one would hope, whereas their magnitudes are comparable. The latter fact suggests that the informational content of the indicators about the true reference group variables is of similar magnitude, but of course not equal. (Otherwise, the indicators would have shown perfect correlation). The standard errors are small, so that the α's have been estimated with considerable accuracy.

Although the estimation results for the factor analysis model are promising, it of course remains to be seen whether the constructed "true variables" are capable of explaining family labour supply better than the indicators used in the previous chapters. This is the topic of the next section.

4.5. The labour supply model reestimated

It should be realized that the true reference group means defined in the previous section cannot be observed directly. Hence, if one wants to use these variables in the labour supply model it is necessary to replace the true variables by proxies. It is well-known however that replacing unobservable variables in a model by proxies leads to inconsistent parameter estimates. Furthermore, the number of observations for which all indicators used in the factor analysis model are available is limited. We will explain in detail how these problems are dealt with in this section. Basically, the solution amounts to the following.

The proxies used for the unobservable true reference group variables are weighted means of the indicators. Given the estimates of the factor analysis model, one can construct these means optimally, in the sense that a proxy gives the best prediction of the corresponding true variable. Furthermore, the estimates of the variances of the ε's in the factor analysis model can be used to compute the inconsistency that would arise in estimation if the proxies were used in estimation of the labour supply model withour further precautions. As a result it is also possible to introduce a correction which guarantees consistent (and actually efficient) estimates of all parameters. To be more precise:

The basic model is given by (4.16)-(4.24) for singles and (4.25)-(4.42) for families. In matrix format the model can be written as

$$x = F\xi + \nu \tag{4.43}$$

The variance covariance matrix of the error vector ν will be denoted by Ψ. In addition to what was mentioned in Section 4.4 we assume that the latent vector ξ is normally distributed with mean zero and variance-covariance matrix Ω. The latter assumption is rather harmless, as we will condition on the values of ξ most of the time.

The vector ξ of latent variables appears in the labour supply equations via the deltas given by (4.1)-(4.2). Apart from the term which contains individual k's own past number of hours, the right hand side of these equations represents the reference group variables of the husband and wife in family k. The latent variables in ξ represent exactly the weighted averages that appear in these equations including reference group family size, which in Section 4.3 was only introduced later. In general terms the delta-equations can be written in matrix format as follows

$$\delta_k = c_k + B\xi_k \tag{4.44}$$

where

$$B=\begin{bmatrix} \omega_{mm} & \omega_{mf} \\ \omega_{fm} & \omega_{ff} \end{bmatrix} \text{ is a matrix of parameters}$$

$$\xi_k=\begin{bmatrix} \xi_{hm}(-1)-\delta_{m1}\xi_{fs}(-1) \\ \xi_{hf}(-1)-\delta_{f1}\xi_{fs}(-1) \end{bmatrix}$$

For families c_k is defined as

$$c_k=\begin{bmatrix} \delta_{m0}+\delta_{m1}f_k+\tau_m[h_{mk}(-1)-\delta_{m1}f_k(-1)] \\ \delta_{f0}+\delta_{f1}f_k+\tau_f[h_{fk}(-1)-\delta_{f1}f_k(-1)] \end{bmatrix} \tag{4.45}$$

For singles $c_k=\delta_0+\delta_1 f_k(-1)+\tau(h_k(-1)-\delta_1 f_k(-1))$ (4.46)

It should be noted that the notation has changed a little relatively to the previous models, because the present model has a simpler specification of the reference group variables. It follows from the factor analysis model that there are two different variables for reference group hours, one for the male and one for the female, whereas there is only one reference group variable for family size. This latter variable is shared by both spouses in a family. The parameters ω indicate the total influence of reference group hours on the δ's of the male and female in a household respectively. For example, ω_{mf} indicates the influence of the reference group mean of female working hours on δ_m. The τ's measure the total influence of habit formation. For the rest the meaning of the parameters is the same as before.

It is convenient to also write the labour supply equations in matrix format (ignoring rationing)

$$h_k = \delta_k + Aw_k + \beta I_k^* + \varepsilon_k \tag{4.47}$$

$$I_k^* = I_k + \vartheta + \delta_k' w_k + 1/2 w_k' A w_k \tag{4.48}$$

where the vector ε_k is supposed to be normally distributed with mean zero and variance covariance matrix Σ. Notice that model (4.43)-(4.48) applies to both the one-and the two-adult household model.

Combining (4.43)-(4.48) leads to the following equation for the vector of household labour supplies

$$h_k = (ID_2 + \beta w_k') c_k + (ID_2 + \beta w_k') B \xi_k + Aw_k + \beta I_k^{**} + \varepsilon_k \tag{4.49}$$

where $I_k^{**} = I_k + \vartheta + 1/2\ w_k' A w_k$

ID_2 = (two by two) identity matrix. For simplicity of notation we write (4.49) as

$$h_k = d_k + D_k \xi_k + \varepsilon_k \tag{4.50}$$

in obvious notation.

The basic statistical model now is (4.43) and (4.50). The first point to note is that not all elements of the vector x are observed for all respondents, although in the estimation of (4.43) only households have been used for whom all elements of x were known.

In the second place the vector of latent variables ξ is not observable, so that we cannot use (4.50) directly. Instead of ξ we observe x. The way to get around the problem of not observing ξ is to construct on the basis of (4.43) and (4.50) an expression of the expectation of h_k conditional on d_k and x_k. In this expression normality of all the variables involved is used heavily. The variance covariance matrix of d_k is denoted by Δ.

Let z_k be the subvector of x_k which is observable for household k. There holds

$$z_k = F_k \xi_k + \nu_k \tag{4.51}$$

where F_k is the part of F corresponding with z_k and ν_k corresponds with z_k as well. The variance covariance matrix of ν_k is denoted by Ψ_k. It is the appropriate submatrix of Ψ. The subscript k has only been added to ξ to indicate that its value will vary across households. It is still the full vector of latent variables.

Using (4.50) and (4.51) we can now write down the expectation of h_k given d_k and z_k. We obtain

$$E(h_k|z_k,d_k)=d_k+D_kE(\xi_k|d_k,z_k)=d_k+ D_k \Omega F_k'(F_k\Omega F_k'+\Psi_k)^{-1}z_k \tag{4.52}$$

$$Var(h_k|z_k,d_k)= \Delta+D_k(\Omega-\Omega F_k'(F_k\Omega F_k'+\Psi_k)^{-1}F_k\Omega)D_k' \tag{4.53}$$

Notice that, after the factor analysis model has been estimated, most of the quantities appearing in (4.52) and (4.53) are known, except the parameters inherent in d_k and D_k.

As said before, for many observations not all indicators of the reference group means are available. One can still construct proxies for the true variables, that are optimal given the amount of information available. Thus, all observations can be used in estimation of the labour supply model. Table 4.12 presents the results. Since the information that is needed for estimation of the factor analysis model is only found in the

SEP questionnaire, we are not able to present OSA-estimation results in this section, unlike in the other chapters of this thesis.

To see whether the use of direct information on reference group variables changes the characteristics of the model, we first compare Table 4.13 with the corresponding columns of Tables 4.1 and 4.2. The values of α, β_m, β_f for families and β and γ for singles are basically the same. The parameters γ_m and γ_f are very different. From Table 4.13 can be concluded that all wage elasticities are lower than in the model with preference formation in which no direct information on reference groups was used (compare Table 4.3). Also the δ_1's are different, but as has been seen before this does not mean too much, as the reduced form coefficients for the influence of family size are rather complicated functions of the structural parameters so that the behaviour of the model is not necessarily very different as a result of the different parameter estimates.

If we now turn to the preference formation parameters, we observe a value for τ_f equal to one, and for τ_m values equal to 0.54 for single males and 0.69 for married males (τ_m and τ_f indicate the totale influence of habit formation). Once again habit formation for females is stronger than for males. Just as in the model described in section 4.2 females are much more referred to than males (ω_{mm} and ω_{fm} are smaller than ω_{ff} and ω_{mf}). It is a bit hard to compare both models, because of the different treatment of reference groups in the two versions. The total effects given in Table 4.14 will shed light on this problem.

The similarities between the total effects of habit formation is clear (compare Table 4.4). As in Section 4.3 we have to conclude that there is a strong effect of mean hours worked by females in the reference group on the δ in the equation. As before we see that preference formation contributes significantly to the explanation of household labour supply, with habit formation dominating the influence of preference interdependence.

There is a striking similarity between the labour supply curves shown in Figures 4.7 and 4.4. This implies that the total effects of reference groups is estimated at about the same magnitude for both the more sophisticated treatment of the reference groups in this chapter and for the more simple specification. This similarity also holds for the hours distributions of married males and females (Figures 4.8 and 4.6).

But for single males and females the simulated hours distributions of the more sophisticated model appear to fit the actual distributions worse.

It should be noted here that although the information used on reference groups is richer than in the preceding chapters, we have not been able to use direct questions on reference group hours, similarly to the questions on reference group family size. The reason for this is that the questionnaire was designed for a project in which hours were not important. The direct questions on reference group incomes, which were included, could not be used. It seems safe to say that direct questions on hours worked in one's reference group would improve the quality of the estimates.

Table 4.12 Estimation results for the extended model, SEP[a)]

	singles			families	
	males				
δ_0	6	(7)	α	-0.05	(0.04)
β	-0.006	(0.004)	β_m	-0.006	(0.001)
δ_1	-0	(1)	β_f	-0.000001	(u.b.)
γ	0.1	(0.3)	γ_m	3.9	(4.2)
ϑ	-1650	(fixed)	γ_f	0.7	(0.2)
τ	0.54	(0.06)	δ_{m1}	-1.7	(0.3)
ω_{mm}	0.4	(1.0)	δ_{f1}	-0.8	(0.5)
ω_{mf}	1	(u.b.)	δ_{m0}	560	(600)
σ_h	12.7	(0.6)	δ_{f0}	-17.0	(2.3)
σ_v	1.10^5	(1.10^6)	ϑ	89027	(97587)
log lik.	-492.1		τ_m	0.69	(0.02)
			τ_f	1	(u.b.)
	females		ω_{mm}	0	(ℓ.b.)
			ω_{mf}	1	(u.b.)
δ_0	-12	(3)	ω_{ff}	1	(u.b.)
β	-0.014	(0.003)	ω_{fm}	0	(ℓ.b.)
δ_1	-0.5	(1.1)	σ_m	7.7	(0.1)
γ	0.7	(0.3)	σ_f	11.3	(0.3)
ϑ	-355	(fixed)	ρ	0.1	(0.1)
τ	1	(u.b.)	σ_{mv}	1238	(1183)
ω_{ff}	1	(u.b.)	σ_{fv}	662	(1562)
ω_{fm}	0	(ℓ.b.)			
σ_h	9.9	(0.5)			
σ_v	438	(223)			
log lik.	-344.2			-4722.6	

a) u.b. = upper bound, ℓ.b. = lower bound
Standard errors in parentheses

Table 4.13 Wage elasticities for the extended model[a)]

	SEP one-adult	SEP two-adult households
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	-0.03	-0.03
	(0.15)	(0.06)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.09	0.48
	(0.22)	(0.51)

a) Standard errors in parentheses

Table 4.14: Total influence of habit formation and preference interdependence

	SEP singles		SEP families	
	influence on			
influence of	δ_{mk}	δ_{fk}	δ_{mk}	δ_{fk}
habit formation, $h_{mk}(-1)$	0.54	-	0.69	-
habit formation, $h_{fk}(-1)$	-	1	-	1
mean of h_m in r.g., ξ_{hm}	0.4	1	0	1
mean of h_f in r.g., ξ_{hf}	1	0	1	0

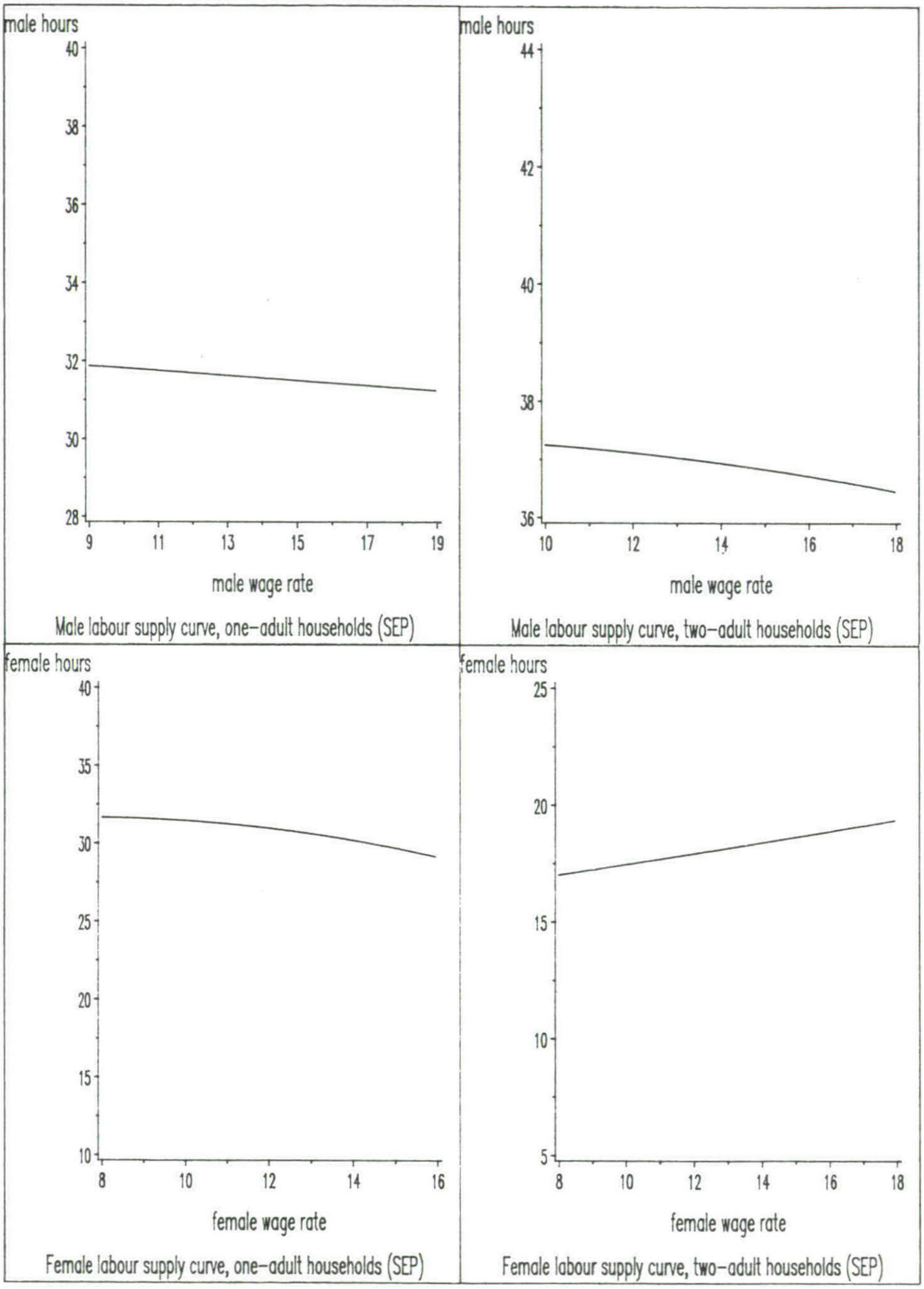


Figure 4.7 <u>Labour supply curves (SEP)</u>

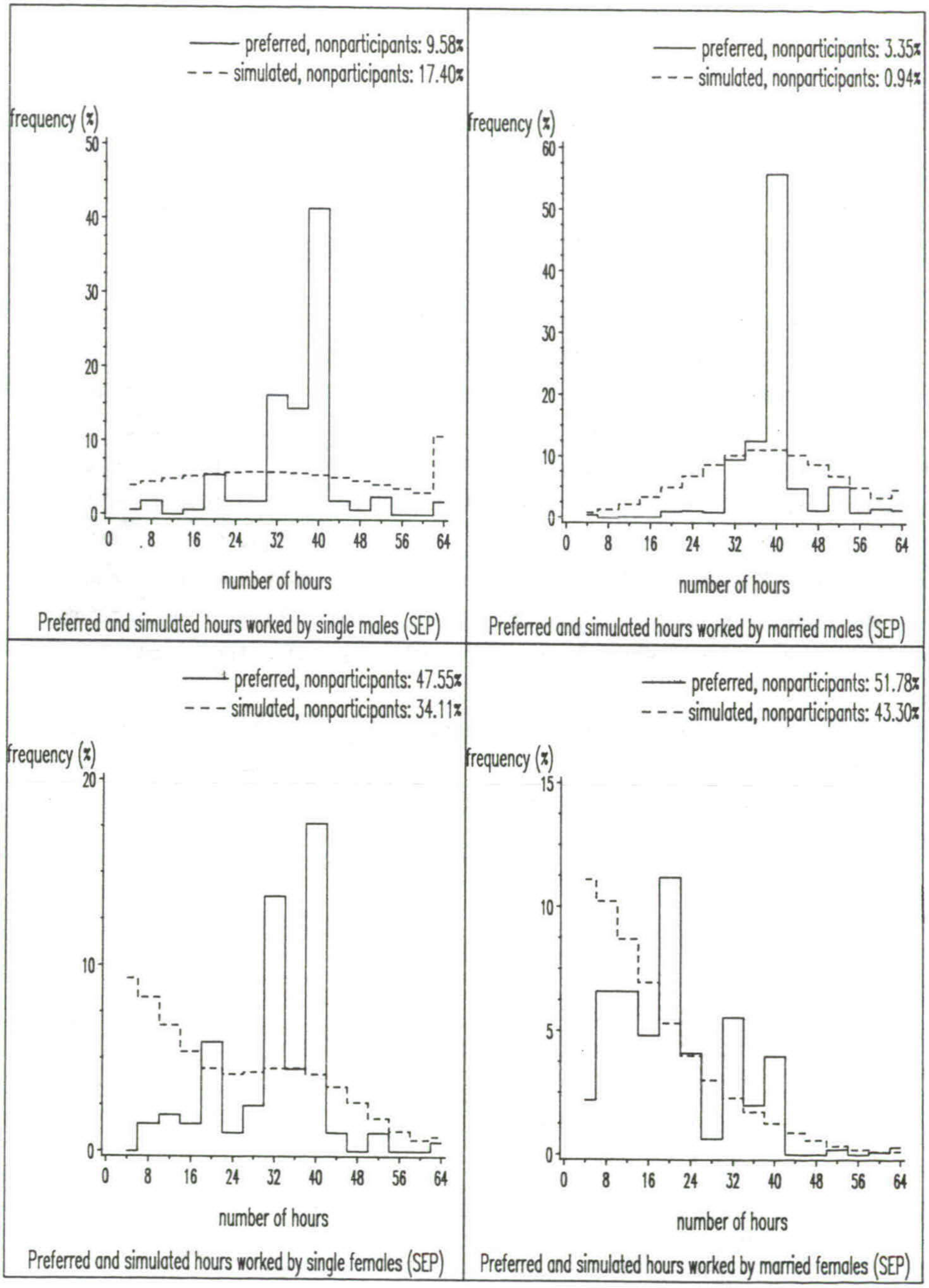


Figure 4.8 Hours distributions (SEP)

4.6 Conclusion

The modelling of preference formation is the major point of emphasis in this chapter. To that end we constructed reference group means of hours worked and family size in Section 4.3 on the basis of a set of assumptions. In Section 4.5 we have employed different pieces of information available in the SEP-data set to construct latent variables which may be thought to be a more accurate reflection of the notion of reference group means than the ones used in Section 4.3

The latent variables model strenghtens the conclusions obtained in Section 4.3 about the determinants of shifts in preferences. For all different groups (single males, single females, families) the estimation results indicate that the incorporation of preference formation leads to different conclusions than would otherwise be obtained. One of the most interesting features of the model is that incorporating preference formation leads to a decrease in the wage elasticities. The statistical significance of preference formation is clearly shown (see Section 4.4 and 4.5).

Since the model is dynamic we are able to distinguish between long term and short term wage effects. Here we have concentrated only on short run effects. For long run effects is referred to Kapteyn, Woittiez and ten Hacken (1989).

Appendix 4A Social group means

Table 4A.1 Social group means, OSA[a)]

	male				female			
	education level[b)]				education level			
	1	2	3	4	1	2	3	4
age 18-30								
log(family size)	0.82	0.84	0.70	0.50	0.92	0.95	0.78	0.57
hours worked, lagged 1 year	32.44	35.06	34.22	29.98	12.74	16.23	20.21	24.83
number of indiv.	43	88	134	58	74	112	235	65
age 30-40								
log(family size)	1.18	1.11	1.16	1.08	1.26	1.28	1.20	0.94
hours worked, lagged 1 year	40.80	38.69	40.50	38.69	7.43	4.90	11.35	18.63
number of indiv.	71	96	281	160	136	146	234	92
age 40-50								
log(family size)	1.28	1.25	1.26	1.28	1.26	1.34	1.28	1.17
hours worked, lagged 1 year	34.11	37.78	39.35	39.69	6.01	4.66	9.90	17.98
number of indiv.	74	85	161	107	112	118	135	50
age 50-65								
log(family size)	0.90	1.02	1.05	0.96	0.96	0.87	0.83	0.64
hours worked, lagged 1 year	20.99	29.56	31.89	34.15	2.95	7.36	9.86	13.08
number of indiv.	136	69	159	93	166	92	84	24

a) 3690 individuals in households and single persons were used to form the social group means.

b) Education has been coded in 4 levels ranging from 1 (lowest) till 4 (highest).

Table 4A.2 Social group means, SEP[a)]

	male				female			
	education level[b)]				education level			
	1	2	3	4	1	2	3	4
age 18-30								
log(family size),	0.97	1.09	0.95	0.73	1.02	1.05	0.77	0.57
hours worked,	24.65	27.08	28.45	32.41	13.47	22.53	23.04	25.19
both lagged one year								
number of indiv.	43	101	152	51	34	72	140	31
age 30-40								
log(family size),	1.18	1.21	1.21	1.23	1.26	1.16	1.13	0.97
hours worked,	28.19	37.78	40.47	40.41	9.21	14.84	17.65	23.61
both lagged one year								
number of indiv.	53	99	243	116	24	62	69	36
age 40-50								
log(family size),	1.20	1.30	1.30	1.32	1.12	1.43	0.92	1.11
hours worked,	29.46	33.80	40.04	40.97	8.50	9.34	22.95	21.45
both lagged one year								
number of indiv.	48	41	147	68	26	32	38	20
age 50-65								
log(family size),	0.83	0.99	0.96	0.95	0.61	0.76	0.79	0.45
hours worked,	17.01	21.96	26.82	34.35	5.97	6.52	9.26	10.71
both lagged one year								
number of indiv.	82	47	106	55	70	33	38	14

a) 2196 individuals in households and single persons were used to form the social group means.

b) Education has been coded in 4 levels ranging from 1 (lowest) till 4 (highest).

Appendix 4B Identification

The model without preference formation is (over)identified. Thus it is sufficient to examine the delta-equations (4.13)-(4.14) for identification of the model with preference formation for families and equation (4.15) for singles.

In Table 4B.1 we have rewritten the delta-equation for single males. As one can see, there are 8 reduced form parameters (a1-a8) and 7 structural parameters (η_{mm}, η_{mf}, ζ_m, κ_m, κ_f, δ_{m1}, δ_{m0}). From a2 δ_{m1} can be identified. Since a5, a6 and a7 can be written as functions of a1, a2, a3 and a4 they do not yield independent information. In fact we have 5 equations (a1, a2, a3, a4 and a8) to identify 7 parameters. By fixing 2 parameters , e.g. κ_m and κ_f the remaining parameters are identified.

Table 4B.2 presents equations (4.13)-(4.14) in a slightly different form. From this table we infer that there are 20 reduced form parameters (a_1-a_{20}) and 16 structural parameters, namely: η_{mm}, η_{ff}, η_{mf}, η_{fm}, λ_{mm}, λ_{ff}, λ_{mf}, λ_{fm}, ζ_m, ζ_f, κ_m, κ_f, δ_{m1}, δ_{f1}, δ_{m0}, δ_{f0}. It is easy to see that δ_{m1} is identified from a2, and likewise δ_{f1} from a12. From a3 and a5 we identify λ_{mm}, from a4 and a6 λ_{mf}, from a13 and a15 λ_{fm}, and from a14 and a16 λ_{ff}. The equations a7, a8 and a9 do not yield independent identifying information, and neither do the equations a17, a18 and a19. For these equations can be written as follows

$$a7 = -a1a2 \tag{4B.1}$$

$$a8 = -a2(a3+a6) \tag{4B.2}$$

$$a9 = -a2(a4+a5) \tag{4B.3}$$

$$a17 = -a11a12 \tag{4B.4}$$

$$a18 = -a12(a13+a16) \tag{4B.5}$$

$$a19 = -a12(a14+a15) \tag{4B.6}$$

In fact we have only 14 independent equations (a1, a2, a3, a4, a5, a6, a10, a11, a12, a13, a14, a15, a16, a20) to identify 16 parameters. If 2 parameters are fixed, the remaining 14 parameters are identified (some combinations of fixed parameters still cause problems, e.g. fixed η_{mm} and ζ_m).

Table 4B.1 The delta-equation for singles

dependent variable δ_m

	structural parameter	reduced form parameter
$h_{mk}(-1)$	$\eta_{mm}\zeta_m$	a1
f_k	δ_{m1}	a2
$\bar{h}_{mmk}(-1)$	$\eta_{mm}(1-\zeta_m)(1-\kappa_m)$	a3
$\bar{h}_{mfk}(-1)$	$\eta_{mf}(1-\kappa_f)$	a4
$\bar{f}_k(-1)$	$-\delta_{m1}\eta_{mm}\zeta_m$	a5
$\bar{f}_{mk}(-1)$	$-\eta_{mm}(1-\zeta_m)(1-\kappa_m)\delta_{m1}$	a6
$\bar{f}_{fk}(-1)$	$-\eta_{mf}(1-\kappa_f)\delta_{m1}$	a7
constant	$\delta_{m0}+\eta_{mm}(1-\zeta_m)\kappa_m(\bar{\bar{h}}_m(-1)-\delta_{m1}\bar{\bar{f}})+\eta_{mf}\kappa_f(\bar{\bar{h}}_f(-1)-\delta_{m1}\bar{\bar{f}})$	a8

Table 4B.2 The delta-equations for families

dependent variable δ_m

	structural parameter	reduced form parameter
$h_{mk}(-1)$	$\eta_{mm}\zeta_m$	a1
f_k	δ_{m1}	a2
$\bar{h}_{mmk}(-1)$	$\eta_{mm}(1-\zeta_m)(1-\kappa_m)\lambda_{mm}$	a3
$\bar{h}_{ffk}(-1)$	$\eta_{mf}(1-\kappa_f)(1-\lambda_{mf})$	a4
$\bar{h}_{fmk}(-1)$	$\eta_{mm}(1-\zeta_m)(1-\kappa_m)(1-\lambda_{mm})$	a5
$\bar{h}_{mfk}(-1)$	$\eta_{mf}(1-\kappa_f)\lambda_{mf}$	a6
$\bar{f}_k(-1)$	$-\delta_{m1}\eta_{mm}\zeta_m$	a7
$\bar{f}_{mk}(-1)$	$-\eta_{mm}(1-\zeta_m)(1-\kappa_m)\lambda_{mm}\delta_{m1}-\eta_{mf}(1-\kappa_f)\lambda_{mf}\delta_{m1}$	a8
$\bar{f}_{fk}(-1)$	$-\eta_{mm}(1-\zeta_m)(1-\kappa_m)(1-\lambda_{mm})\delta_{m1}-\eta_{mf}(1-\kappa_f)(1-\lambda_{mf})\delta_{m1}$	a9
constant	$\delta_{m0}+\eta_{mm}(1-\zeta_m)\kappa_m(\bar{\bar{h}}_m(-1)-\delta_{m1}\bar{\bar{f}})+\eta_{mf}\kappa_f(\bar{\bar{h}}_f(-1)-\delta_{m1}\bar{\bar{f}})$	a10

dependent variable δ_f

	structural parameter	reduced form parameter
$h_{fk}(-1)$	$\eta_{ff}\zeta_f$	a11
f_k	δ_{f1}	a12
$\bar{h}_{mmk}(-1)$	$\eta_{fm}(1-\kappa_m)(1-\lambda_{fm})$	a13
$\bar{h}_{ffk}(-1)$	$\eta_{ff}(1-\zeta_f)(1-\kappa_f)\lambda_{ff}$	a14
$\bar{h}_{fmk}(-1)$	$\eta_{fm}(1-\kappa_m)\lambda_{fm}$	a15
$\bar{h}_{mfk}(-1)$	$\eta_{ff}(1-\zeta_f)(1-\kappa_f)(1-\lambda_{ff})$	a16
$\bar{f}_k(-1)$	$-\delta_{f1}\eta_{ff}\zeta_f$	a17
$\bar{f}_{mk}(-1)$	$-\eta_{ff}(1-\zeta_f)(1-\kappa_f)(1-\lambda_{ff})\delta_{f1}-\eta_{fm}(1-\kappa_f)(1-\lambda_{fm})\delta_{f1}$	a18
$\bar{f}_{fk}(-1)$	$-\eta_{ff}(1-\zeta_f)(1-\kappa_f)\lambda_{ff}\delta_{f1}-\eta_{fm}(1-\kappa_m)\lambda_{fm})\delta_{f1}$	a19
constant	$\delta_{f0}+\eta_{ff}(1-\zeta_f)\kappa_f(\bar{\bar{h}}_f(-1)-\delta_{f1}\bar{\bar{f}})+\eta_{fm}\kappa_m(\bar{\bar{h}}_m(-1)-\delta_{f1}\bar{\bar{f}})$	a20

Appendix 4C Stability

To investigate the stability of the two-adult household model we write the equations (3.7)-(3.12) and (4.13)-(4.14) in matrix notation

$$h = \left[\beta \otimes W + ID_{2K}\right]\delta + \sum_{j=1}^{7}\left[\gamma_j \otimes x_j\right] + \varepsilon \tag{4C.1}$$

$$\delta = \delta_0 \otimes \iota_K + \left[E_1\ Z \otimes ID_K\right]\ h(-1) + \left\{\left[E_1\left[I_2 - Z\right] \otimes ID_K\right]Q_1 + \right.$$
$$\left. + \left[E_2 \otimes ID_K\right]Q_2\right\}\ h(-1) \tag{4C.2}$$

where $h = \left[h_{m1}\ \cdots\ h_{mK} h_{f1}\ \cdots\ h_{fK}\right]'$

K = total number of households

$$\beta = \begin{bmatrix}\beta_m \\ \beta_f\end{bmatrix}$$

$$W = \begin{bmatrix} w_{m1} & & 0 & w_{f1} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & w_{mK} & 0 & & w_{fK}\end{bmatrix}$$

$$\delta = \left[\delta_{m1}\ \cdots\ \delta_{mK}\ \delta_{f1}\ \cdots\ \delta_{fK}\right]'$$

$$\gamma_1 = \begin{bmatrix}\gamma_m \\ \alpha\end{bmatrix} \quad \gamma_2 = \tfrac{1}{2}\begin{bmatrix}\beta_m\gamma_m \\ \beta_f\gamma_m\end{bmatrix} \quad \gamma_3 = \begin{bmatrix}\alpha \\ \gamma_f\end{bmatrix} \quad \gamma_4 = \tfrac{1}{2}\begin{bmatrix}\beta_m\gamma_f \\ \beta_f\gamma_f\end{bmatrix}$$

$$\gamma_5 = \begin{bmatrix}\beta_m\alpha \\ \beta_f\alpha\end{bmatrix} \quad \gamma_6 = \begin{bmatrix}\beta_m \\ \beta_f\end{bmatrix} \quad \gamma_7 = \begin{bmatrix}\beta_m\vartheta \\ \beta_f\vartheta\end{bmatrix}$$

$$x_1 = \left[w_{m1}\ \cdots\ w_{mK}\right]'$$

$$x_2 = \left[w_{m1}^2\ \cdots\ w_{mK}^2\right]'$$

$$x_3 = \left[w_{f1}\ \cdots\ w_{fK}\right]'$$

$$x_4 = \left[w_{f1}^2\ \cdots\ w_{fK}^2\right]'$$

$$x_5 = \left[w_{m1} w_{f1} \cdots w_{mK} w_{fK}\right]'$$

$$x_6 = \left[I_1 \cdots I_K\right]'$$

$$x_7 = (1 \ldots 1)' \triangleq \iota_K'$$

$$\delta_0 = \begin{bmatrix} \delta_{m0} \\ \delta_{f0} \end{bmatrix}$$

$$E_1 = \begin{bmatrix} \eta_{mm} & 0 \\ 0 & \eta_{ff} \end{bmatrix}; \; E_2 = \begin{bmatrix} \eta_{mf} & 0 \\ 0 & \eta_{fm} \end{bmatrix}$$

$$Z = \begin{bmatrix} \zeta_m & 0 \\ 0 & \zeta_f \end{bmatrix}$$

$$Q^{ii} = \begin{bmatrix} 0 & q_{12}^{ii} & \cdots & q_{1K}^{ii} \\ q_{21}^{ii} & \ddots & & \vdots \\ \vdots & & \ddots & q_{K-1,K}^{ii} \\ q_{K1}^{ii} & \cdots & q_{K,K-1}^{ii} & 0 \end{bmatrix}, \quad Q^{ij} = \begin{bmatrix} q_{11}^{ij} & \cdots & q_{1K}^{ij} \\ \vdots & \ddots & \vdots \\ q_{Ki}^{ij} & \cdots & q_{KK}^{ij} \end{bmatrix}$$

$$Q_1 = \begin{bmatrix} Q^{mm} & 0 \\ 0 & Q^{ff} \end{bmatrix} \qquad Q_2 = \begin{bmatrix} 0 & Q^{mf} \\ Q^{fm} & 0 \end{bmatrix}$$

where $i,j = m,f$

ID_K = (K by K) identity matrix.

It is assumed that in the long run preferred number of hours, actual number of hours and actual number of hours one period lagged are equal. Substituting (4C.2) into (4C.1) then yields

$$h = \left\{ID_{2K} - G\left[E_1 Z \otimes ID_K + \left[E_1\left(I_2 - Z\right) \otimes ID_K\right] Q_1 + \left[E_2 \otimes ID_K\right] Q_2\right]\right\}^{-1}$$
$$\left\{G\left[\delta_o \otimes \iota_N\right] + \sum_{i=1}^{7}\left(\gamma_i \otimes x_i\right) + \varepsilon\right\} \tag{4C.3}$$

where $G = \beta \otimes W + ID_{2K}$

Define $H = G\left[E_1 Z \otimes ID_K + \left[E_1\left(ID_2 - Z\right) \otimes ID_K\right] Q_1 + \left[E_2 \otimes ID_K\right] Q_2\right]$ (4C.4)

Stability of the model is guaranteed when it can be shown that the eigenvalues of H lie within the unit circle. By Gershgorin's Theorem (see Marcus and Minc (1964)) we are able to specify an upper and lower bound for the eigenvalues.

Gershgorin's Theorem. Each eigenvalue of the K x K matrix $A = [a_{ij}]$ lies in some interval $I_i = [a_{ii} - sa_i, a_{ii} + sa_i]$, $i=1,\dots,K$ where $sa_i = \sum_{j \neq i} a_{ij}$

The investigation of the stability of the one-adult-households is done analagously. For the model described in Section 4.4 the only necessary adjustment is substituting

$$\left[E_1 Z \otimes ID_K\right] \text{ by } \begin{pmatrix} \tau_m & 0 \\ 0 & \tau_f \end{pmatrix}$$

$$\left[E_1\left(ID_2 - Z\right) \otimes ID_K\right] Q_1 \text{ by } \begin{pmatrix} \omega_{mm} & 0 \\ 0 & \omega_{ff} \end{pmatrix}$$

$$\left[E_2 \otimes ID_K\right] Q_2 \text{ by } \begin{pmatrix} \omega_{mf} & 0 \\ 0 & \omega_{fm} \end{pmatrix}$$

By applying Gershgorin's Theorem to the 2K x 2K matrix H and taking sample means for the wage rates, we obtain the union of all 2K intervals for the estimated parameters, presented in Tables 4.1, 4.2 and 4.12

	OSA	SEP	SEP(latent var.)
single males	(0.20, 0.20)	(0.25, 0.79)	(-0.29, 1.3)
single females	(-0.79, 1.42)	(0.84, 0.84)	(0.84, 0.84)
males in families	(-0.02, 0.41)	(-0.07, 1.49)	(0.40, 0.84)
females in families	(0.95, 0.99)	(0.99, 0.99)	(0.99, 0.99)

Thus, for single females in the OSA sample and for males in families in the SEP sample stability is not guaranteed in the model with preference formation, described in Section 4.2. In the latent variable model stability cannot be guaranteed for single males. For all other groups the sufficient conditions for stability are guaranteed.

Appendix 4D Correlation matrices of indicators

Table 4D.1 Correlation matrices of indicators

HOURS

	single males				single females			
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
x_1	1	0.46	0.36	0.31	1	0.41	0.53	0.26
x_2		1	0.22	0.63		1	0.30	0.58
x_3			1	0.33			1	0.52
x_4				1				1

	males in families				females in families			
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4
x_1	1	0.30	0.52	0.37	0.06	0.02	0.13	0.22
x_2		1	0.30	0.57	0.02	0.33	0.11	0.32
x_3			1	0.38	0.10	0.09	0.04	0.09
x_4				1	0.06	0.47	0.09	0.44
x_1					1	0.37	0.49	0.38
x_2						1	0.33	0.35
x_3							1	0.87
x_4								1

Table 4D.1 Correlation matrices of indicators, continued

FAMILY SIZE

	single males					single females				
	x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
x_1	1	0.21	0.88	0.23	0.05	1	0.29	0.89	0.47	0.18
x_2		1	0.15	0.05	0.39		1	0.31	0.74	0.20
x_3			1	0.21	0.05			1	0.48	0.18
x_4				1	-0.18				1	0.21
x_5					1					1

	males in families					females in families				
	x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
x_1	1	0.30	0.54	0.38	0.35	0.56	0.26	0.79	0.27	0.28
x_2		1	0.33	0.55	0.38	0.27	0.41	0.38	0.39	0.23
x_3			1	0.44	0.27	0.74	0.28	0.47	0.29	0.26
x_4				1	0.23	0.42	0.69	0.43	0.68	0.22
x_5					1	0.25	0.10	0.36	0.13	0.54
x_1						1	0.35	0.58	0.36	0.25
x_2							1	0.35	0.90	0.07
x_3								1	0.36	0.25
x_4									1	0.03
x_5										1

5 Job characteristics

5.1 Introduction

The major objective of this chapter is to develop a model of job choice, wages and labour supply, and to use this model to estimate the magnitude of the compensating wage differentials for qualitative differences in jobs (such as pleasantness, riskiness, dirtiness etc.). Like in the preceding chapters, we will maintain a utility maximizing framework as the starting point of the analysis. But we abstain from the idea that the household is a joint decision unit. Neither do we take into account preference formation. Both extensions would complicate the model too much.

In this chapter it is assumed that the individual maximizes a utility function with leisure, nonpecuniary job characteristics and consumption as its arguments. Nonpecuniary job characteristics are defined as nonmonetary benefits accruing to individuals as a consequence of their particular occupation (e.g. working conditions).

Moreover one's wage is assumed to depend on human capital variables and the characteristics of the job actually chosen. The idea behind this relation is that the rewards for working can be both pecuniary (money wage) and nonpecuniary (compensation by means of the supply of desirable job characteristics).

Most empirical work on compensating wage differentials has been based on the estimation of a single wage equation (cf. Hamermesh (1977), Barron and Loewenstein (1986), Antos and Rosen (1975), Duncan (1976)). One major disadvantage of these studies is that information on labour supply is neglected. Since we believe that the labour supply decision is taken jointly with the choice of a particular job with its associated wage, it is essential that this information is incorporated in the analysis. Moreover analyses that ignore the (self) selection of individuals into different occupations, may produce misleading results. Only in a few papers a model is developed in which labour supply, job choice and the determination of wages are analysed jointly. Examples are Atrostic (1982) and Killingsworth (1984). Killingsworth developed a model of labour supply and discrete job choice, and used it in an empirical analysis of the compensating wage differential between white and blue collar work.

Jobs with qualitative differences can be interpreted as differentiated products, described by a vector of objectively measured characteristics. The observed product price (this is the wage rate if we are dealing with the labour market) and the specific amounts of characteristics define a set of implicit or hedonic prices (in our case the prices of job characteristics). Some theoretical papers on hedonic prices are by Rosen (1974), Brown and Rosen (1982).

By including job characteristics in the wage equation, the wage rate becomes an endogenous variable and this leads to a nonlinear budget constraint.

In Section 5.2 a simultaneous model of job choice, labour supply and wages is developed. In Section 5.3 the estimation results are presented. Finally, in Section 5.4 some concluding remarks are made.

5.2 The model

We adopt the home production approach to the theory of consumption, which takes goods and services as inputs in a production process that generates utility bearing outputs usually called commodities (cf. Becker (1965), Lancaster (1966), Gronau (1986)). Let Z_i be the quantity of commodity i and let the production of each commodity require a combination of time inputs $T_j (j=1,..,p)$ and goods $X_\ell (\ell=1,..,n)$

$$Z_i = f_i(X_1^i,..,X_n^i,T_1^i ,..,T_p^i), \quad i=1,..,m \tag{5.1}$$

Here X_ℓ^i is the amount of the ℓ-th good used in the production of the i-th commodity and T_j^i is the amount of the j-th time input used in the production of the i-th commodity; f_i is the production function of the i-th commodity. The utility function is

$$\Phi(Z_1,..,Z_m) \tag{5.2}$$

The total number of goods and the total number of time inputs used in the production of all commodities jointly are respectively

$$X_\ell = \sum_i X_\ell^i, \quad \ell=1,..,n \tag{5.3}$$

$$T_j = \sum_i T_j^i \ , \quad j=1,..,p \tag{5.4}$$

In accordance with previous chapters let unearned income of the individual (or household) be equal to I and assume for simplicity that the individual can hold only one paid job; the amount of time spent on the job is T_p. Restrictions on behaviour are

$$\sum_{\ell} p_{\ell} X_{\ell} = I + w(q)T_p \tag{5.5}$$

where $w(q)$ is the wage rate and

$$\sum_j T_j = T \tag{5.6}$$

where T is the total time endowment.
The wage rate corresponding to a job depends on a vector q of job characteristics.

In principle, behaviour follows from maximization of (5.2) subject to (5.1), (5.3), (5.4), (5.5) and (5.6). It has been noted by various authors (e.g. Pollak and Wachter (1975)) that without further assumptions and without information about commodities, this model cannot be distinguished from a model in which utility is defined directly in terms of goods and time inputs. Examples of papers with specific implications are Gronau (1977,1980) and Graham and Green (1984).

In this paper we consider a static model in which relative prices of consumption goods are constant across consumers. Furthermore, the structure of the production functions f_i is as follows

$$Z_i = f_i(X_1^i,..,X_n^i,T_1^i,..,T_p^i, q), \quad i=1,..,m \tag{5.7}$$

The vector q appears in the production function because we want to allow for the possibility that at least some commodities are being influenced by the type of the job one is holding. This means that q is not taken as given, but it is modelled as a choice variable. An example might be that a commodity is the amenity of a job and that q describes the occupational status of the job.

Given that relative prices of consumption goods are assumed to be constant across consumers, perfect aggregation over goods is possible. Similarly, we can aggregate the time inputs $T_1,..,T_{p-1}$ perfectly. Denote the aggregate of the consumer goods by c, the aggregate of the first p-1 time inputs by L and let $h = T_p$. If we are only interested in three commodities, namely leisure (Z_1), the amenity of a job (Z_2) and consumption (Z_3), we can then, without loss of generality, restrict our attention to the following maximization problem

$$\max.\ U(Z_1,Z_2,Z_3) \tag{5.8}$$

$$\text{s.t. } p_c c = w(q)(T-L) + I \tag{5.9}$$

$$Z_i = f_i(L,q,c), \quad i=1,2,3 \tag{5.10}$$

It is important to keep in mind the distinction between goods (inputs in the production process) and commodities (utility bearing outputs). During the following discussion commodities are denoted by Z's and goods by the usual symbols h,q,c. We assume that the production functions f_1 and f_3 are trivial, that is

$$Z_1 = T-L = h \tag{5.11}$$

(We write utility as a function of h (hours of work) instead of L (leisure) to keep in line with the specification used in Chapter 2.)

$$Z_3 = c \tag{5.12}$$

For f_2 we have examined two different specifications, namely

$$Z_2 = hq \tag{5.13}$$

$$Z_2 = q \tag{5.14}$$

Discussion of the interpretation of both specifications is postponed to the appropriate subsections. For the function $w(q)$ we adopt the following linear specification

$$w = a'k + bq \qquad (5.15)$$

where k = vector of individual characteristics
a, b are parameters

The variable q is assumed to be a continuous (latent) variable. Equation (5.15) reflects the fact that the wage rate depends on individual characteristics, such as age and education, as well as on job characteristics such as working conditions. If q is defined as a "good", the expected sign of the compensating wage differential b is negative. For jobs that offer desirable job characteristics can attract labour at a lower than average wage rate, while jobs that offer undesirable job characteristics must pay a premium to attract labour. In fact employees sell the services of their labour, but simultaneously purchase utility bearing job characteristics. Employers purchase the services of labour, and jointly sell job characteristics. The observed relation between wages, human capital of the employee and job characteristics is determined by the market in such a way that employees and employers are matched correctly. For example, assume that individuals have homogeneous preferences for different kinds of work, say all individuals dislike dirty work. Furthermore assume there is a continuum of jobs, then if there is no wage differential between dirty and clean jobs, nobody will be working in dirty jobs. So individuals working in jobs with undesirable characteristics have to be paid extra. This is in short the idea of compensating wage differentials.

So far we have assumed that q is a choice variable, but we can also take q as given, which is common practice in empirical studies on wage differentials. Then we obtain a model to which we will refer as standard model, resulting from the following maximization problem

$$\max \quad U(Z_1, Z_2)$$

$$\text{s.t.} \quad p_c c = w(q)h + I$$

$$Z_1 = h$$
$$Z_2 = c$$

The standard model is comparable to the one-adult household model in Chapter 2. In this chapter we focus, however, on the extended model in which job choice is endogenous. In the following subsections the implications of the two different specifications of the production function of Z_2 in the extended model will be made explicit.

Specification Z_2 = hq

The utility maximization problem can be written as follows

$$\max \quad U(Z_1, Z_2, Z_3) \tag{5.16}$$

$$\text{s.t.} \quad Z_1 = h \tag{5.17}$$
$$Z_2 = hq \tag{5.18}$$
$$Z_3 = c \tag{5.19}$$
$$p_c c = wh + I \tag{5.20}$$
$$w = a'k + bq \tag{5.21}$$

The output level for commodity Z_2 (job amenity) does not only depend on the type of job one is working in with associated characteristic q, but also on the number of hours one is working. The interpretation of this specification is the following: The disutility bearing output resulting from working only a few hours in a tedious job is less than the disutility bearing output of working many hours on that same job, and vice versa for utility bearing outputs. In this interpretation both Z_2 and q are closely connected with time spent on the job. If one doesn't work, then q no longer affects utility. As we will see, this is different in the second specification. Hours of work affect utility through two channels. Assuming that both Z_2 and q are valued positively, more hours of work yield more utility, if one is working in a pleasant job (i.e. if $q>0$). On the other hand, more hours of work means less leisure time and thus less utility. If q is less than zero, that is, one is working in an unpleasant job, both effects work in the same direction.

Following the approach taken by Pollak and Wachter (1975), we define a cost function $C(P,Z)$ as the cost of the least expensive collection of goods, capable of producing the commodity vector $Z=(Z_1,Z_2,Z_3)$ when good prices are P

$$C(P,Z) = \min\ (p_c Z_3 - a'kZ_1 - bZ_2) \tag{5.22}$$

Implicit commodity prices $\pi=(\pi_1,\pi_2,\pi_3)$ are defined as the marginal costs of producing commodities:

$$\pi_1(P,Z) = \frac{\partial C(P,Z)}{\partial Z_1} = -a'k \tag{5.23}$$

$$\pi_2(P,Z) = \frac{\partial C(P,Z)}{\partial Z_2} = -b \tag{5.24}$$

$$\pi_3(P,Z) = \frac{\partial C(P,Z)}{\partial Z_3} = p_c \tag{5.25}$$

Further, implicit income (IC) is defined as the cost of the commodity bundle Z, evaluated at the implicit commodity prices

$$IC\ (= \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3) = I \tag{5.26}$$

Maximization problem (5.16)-(5.21) can be rewritten as

$$\begin{aligned} &\max \quad U(Z_1,Z_2,Z_3) \\ &\text{s.t.} \quad \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 = IC \end{aligned} \tag{5.27}$$

Since the commodity prices are independent of the commodity bundle, chosen by the individual, an analogy with traditional demand theory is preserved. Thus, we can proceed with the specification of the indirect utility function.

For our specification of the indirect utility function we adopt a variant of the Hausman-Ruud model (cf. Hausman and Ruud (1984))

$$V(\pi, IC) = \exp((\beta_1\pi_1 + \beta_2\pi_2)/\pi_3)I^* \tag{5.28}$$

$$I^* = IC/\pi_3+\vartheta+\delta_1\pi_1/\pi_3+\delta_2\pi_2/\pi_3+1/2\gamma_1\pi_1^2/\pi_3^2+1/2\gamma_2\pi_2^2/\pi_3^2 \tag{5.29}$$

Application of Roy's Identity yields demand functions for the commodities Z

$$Z_1 = \beta_1 I^* + \delta_1 + \gamma_1\pi_1/\pi_3 \tag{5.30}$$

$$Z_2 = -\beta_2 I^* - \delta_2 - \gamma_2\pi_2/\pi_3 \tag{5.31}$$

$$Z_3 = (\beta_1\pi_1/\pi_3+\beta_2\pi_2/\pi_3+1)I^* -\vartheta + 1/2\gamma_1\pi_1^2/\pi_3^2 + 1/2\gamma_2\pi_2^2/\pi_3^2 \tag{5.32}$$

The price of consumption π_3 is used as a numeraire. Translating system (5.30)-(5.32) back into the goods-space, the following goods demand functions are derived

$$h = \beta_1\bar{I} + \delta_1 + \gamma_1 a'k \tag{5.33}$$

$$hq = -\beta_2\bar{I} - \delta_2 + \gamma_2 b \tag{5.34}$$

$$c = (\beta_1 a'k-\beta_2 b +1)\bar{I} -\vartheta + 1/2\gamma_1(a'k)^2 + 1/2\gamma_2 b^2 \tag{5.35}$$

where $$\bar{I} = I+\vartheta+\delta_1 a'k-\delta_2 b+1/2\gamma_1(a'k)^2+1/2\gamma_2 b^2 \tag{5.36}$$

Equation (5.33) corresponds with the demand equations, derived in Chapter 2 (equation (2.1)). One difference with the model in Chapter 2, however, is that in this chapter the wage rate is modelled jointly with the labour supply decision, while in Chapter 2 it was modelled seperately and implemented afterwards into the labour supply equation. Equation (5.34) is the demand equation for job amenities, while equation (5.35) is the demand equation for consumption.

The price of commodity Z_1 (hours of work) is parameterized as follows

$$a'k = a_1 + a_{22}educ2 + a_{23}educ3 + a_{24}educ4 + a_{31}age + \tag{5.37}$$

$$a_{32}\,age^2 + a_{41}edsec1 + a_{42}edsec2 + a_5supvis + a_6private$$

where edsec1 = 1 if one's education is technical

= 0 if it is not

edsec2 = 1 if one's education is in economics

0 if it is not

supvis = 1 if one is a supervisor of more than 10 persons

0 if one is not

private = 1 if one is working in a private sector

0 if one is not

Different educational sectors are included in the wage equation to investigate whether the sector one is educated in influences one's wage rate. To deal with demographic variation we have parameterized δ_1 as

$$\delta_1 = \delta_{10} + \delta_{11}f \tag{5.38}$$

where f = log of family size.

The parameter δ_2 is parameterized as follows

$$\delta_2 = \delta_{21}age + \delta_{22}educ2 + \delta_{23}educ3 + \delta_{24}educ4 \tag{5.39}$$

We would expect that the higher one's age and the higher one's education, the stronger one's preferences for job amenities. For definition of the variables is referred to Chapter 2. Notice that the system of goods demand functions described above is nonlinear; the nonlinearity comes in through equation (5.34). In estimation we will use the reduced form of the model, which is easy to derive (see Appendix 5A). Before turning to the stochastic version of the model, we will first describe the implications of the second specification of the production function.

Specification $Z_2 = q$

The utility maximization problem has the following form

$$\max \quad U(Z_1, Z_2, Z_3) \tag{5.40}$$

$$\text{s.t.} \quad Z_1 = h \tag{5.41}$$

$$Z_2 = q \tag{5.42}$$

$$Z_3 = c \tag{5.43}$$

$$p_c c = wh + I \tag{5.44}$$

$$w = a'k + bq \tag{5.45}$$

Following the same procedure as described for the previous specification, a cost function is specified

$$C(P,Z) = \min \; (p_c Z_3 - a'kZ_1 - bZ_1Z_2) \tag{5.46}$$

with good prices P.
From this cost function implicit prices of the commodities are derived

$$\pi_1(P,Z) = \frac{\partial C(P,Z)}{\partial Z_1} = -a'k - bZ_2 \tag{5.47}$$

$$\pi_2(P,Z) = \frac{\partial C(P,Z)}{\partial Z_2} = -bZ_1 \tag{5.48}$$

$$\pi_3(P,Z) = \frac{\partial C(P,Z)}{\partial Z_3} = p_c \tag{5.49}$$

Price π_3 is used as a numeraire. And implicit income (IC) is defined as

$$IC(P,Z) \; (= \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3) = I - bZ_1Z_2 \tag{5.50}$$

Since both implicit prices and implicit income depend on the commodity bundle consumed, the application of traditional demand theory is not straightforward. In fact, Pollak and Wachter (1975) argue that in this case the link between the household production approach and neoclassical theory is broken. In a comment on this paper, Barnett (1977) shows this need not be true. The household decision problem (5.40)-(5.45) is equivalent with

$$\begin{aligned} &\max \quad U(Z_1, Z_2, Z_3) \\ &\text{s.t.} \quad \pi_1 Z_1 + \pi_2 Z_2 + \pi_3 Z_3 = IC(P, Z) \end{aligned} \tag{5.51}$$

where π_1, π_2, π_3 and IC are defined above. Assume that the household's solution value for Z is Z^*, where Z^* is a function of goods prices P and nonlabour income I. Define π^* by $\pi^*(P, Z^*)$ and IC^* by $IC^*(P, Z^*)$. Then Barnett shows that the household acts as if it were solving

$$\begin{aligned} &\max \quad U(Z_1, Z_2, Z_3) \\ &\text{s.t.} \quad \pi_1^* Z_1 + \pi_2^* Z_2 + \pi_3^* Z_3 = IC^*(P, Z) \end{aligned} \tag{5.52}$$

The π^*'s can be interpreted as shadow prices, constructed at the solution point. Since π^* and IC^* do not depend on Z, we can now derive demand functions with the known neoclassical properties.

Before we are turning to the indirect utility function, let us first have a closer look at the maximization problem. The utility function U(h,q,c) is maximized subject to the budget constraint $p_c c = I + w(q)h$. The necessary conditions for a maximum are

$$w/p_c = -\frac{\partial U}{\partial h} \Big/ \frac{\partial U}{\partial c} \tag{5.53}$$

$$h/p_c\left(\frac{\partial w}{\partial q}\right) = -\frac{\partial U}{\partial q} \Big/ \frac{\partial U}{\partial c} \tag{5.54}$$

where, by (5.45), (5.54) can be written as

$$hb\ /p_c = -\frac{\partial U}{\partial q} \Big/ \frac{\partial U}{\partial c} \tag{5.55}$$

The left hand side of (5.55) is the derivative of the earnings function with respect to a job characteristic, and the right hand side is the marginal rate of substitution between a job characteristic and money. Furthermore, if q is defined as a "good", it follows from (5.55) that under reasonable conditions b must be negative. One can also see immediately that for the corner solution h=0 the existence of a satiation point of q is required. This can also be seen in the Figures 5.1 to 5.3, in which the shape of the budget set is shown for different values of b, in respectively the (c,L)-plane, the (c,q)-plane and the (q,L)-plane, where $a'k>0$ and $p_c=1$.

According to (5.44) the budget constraint is a straight line in the (c,L)-plane (see Figure 5.1). If b is less than zero, then more of q leads to a lower wage rate (cf. 5.45), and consequently to less consumption. For q large enough ($q>-\frac{a'k}{b}$) individuals have to pay to work ($w<0$) (See Figure 5.1, the case $b<0$). This can only lead to a well defined optimum if more leisure leads to less utility, since the optimum point now has to be on the rising part of the indifference curve. That is, $\frac{dc}{dL} = -\frac{\partial U}{\partial L}/\frac{\partial U}{\partial c}>0$. If we assume that $\frac{\partial U}{\partial c}>0$, which is always the case in the specification we use, then $\frac{\partial U}{\partial L}<0$. If we assume that there are also points for which $\frac{\partial U}{\partial L}>0$, then this implies the existence of a satiation point for leisure, i.e. $\frac{\partial U}{\partial L}=0$. For $q=-\frac{a'k}{b}$ the budget line is horizontal, so that (given q) the optimal point is the satiation point. Notice that the corner solution h=0 need not be treated as a special case. If $b>0$, more q yields a higher wage rate and thus a higher budget. For the rest, the figure is similar to the figure corresponding with $b<0$. For b=0, for all different values of q the same budget line holds.

In the (c,q)-plane the budget line is also a straight line (cf. 5-.45), but the slope depends on the sign of b (see Figure 5.2). If $b<0$ the slope of the budget line is negative, and increases (in absolute value) if h increases. For h=0 or b=0 the budget line is horizontal and hence an optimum can only arise in a point of complete satiation with respect to q. If $b>0$ the existence of an optimum also implies satiation in some point (but not in the optimum point).

In the (q,L)-plane the budget line is a hyperbola, of which one of the asymptotes is L=T and the other one $q=-\frac{a'k}{b}$ (See Figure 5.3). We distinguish two different situations, namely $c>I$ and $c=I$. If $c>I$ and $b<0$ the budget line has the form as shown in the top figure. Given the number of hours someone works, he can consume less q if he consumes more c. Notice that in this case the budget set is convex. For $c\rightarrow I$ the budget line coincides with the asymptotes. Three different situations are possible in that case: 1) h=0 and $q = -\frac{a'k}{b}$ (point A), then no restrictions are required on preferences, 2) h=0 and $q \neq -\frac{a'k}{b}$, in this case there must be satiation of q (point C), 3) $h \neq 0$ and $q = -\frac{a'k}{b}$, in this case there must be satiation of L (point B). If b>0 (middle figure) then for an optimum it is necessary that in some point there is complete satiation of q. And the same holds for the case b=0.

Summarizing, we can say that for b<0 only in special cases, namely when the wage rate is less than zero or when $q = -\frac{a'k}{b}$, it is required that there exists a point of satiation in leisure. For b>0, there always must be a point of satiation in q, just as in the case of b=0. A rather special case is the corner solution h=0. Then there must exist a satiation point of q (for any value of b).

Like in the standard model in which the budget constraint is a straight line, the necessary condition for utility maximization is that the slope of the indifference curve equals the slope of the budget constraint. But, unlike in the standard model, it is not sufficient for a utility maximum that the indifference curves are convex. Rather, the more stringent assumption is required that the indifference curve is 'more convex' than the budget curve (in Appendix 5A the requirements are given explicitly.). From Figures 5.1 to 5.3 it is clear that the budget line is concave in the (q,L)-plane and a straight line in the two other planes. So we expect that the requirement that the indifference curve is more convex than the budget line is fulfilled. In estimation this condition is checked.

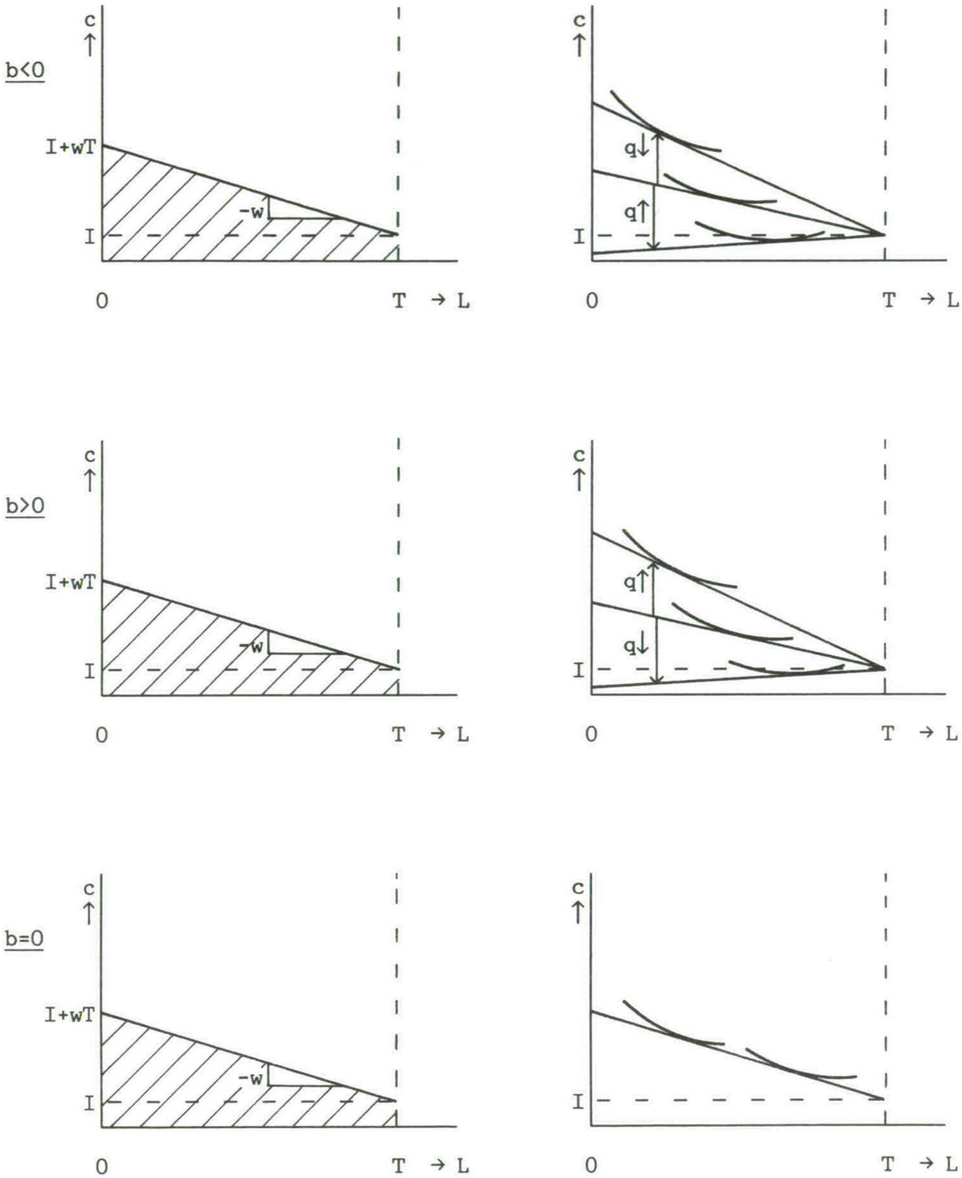


Figure 5.1 <u>The budget set in the (c,L) -plane for a given value of q</u> <u>The optimum situation in the (c,L)-plane for different values of q</u>

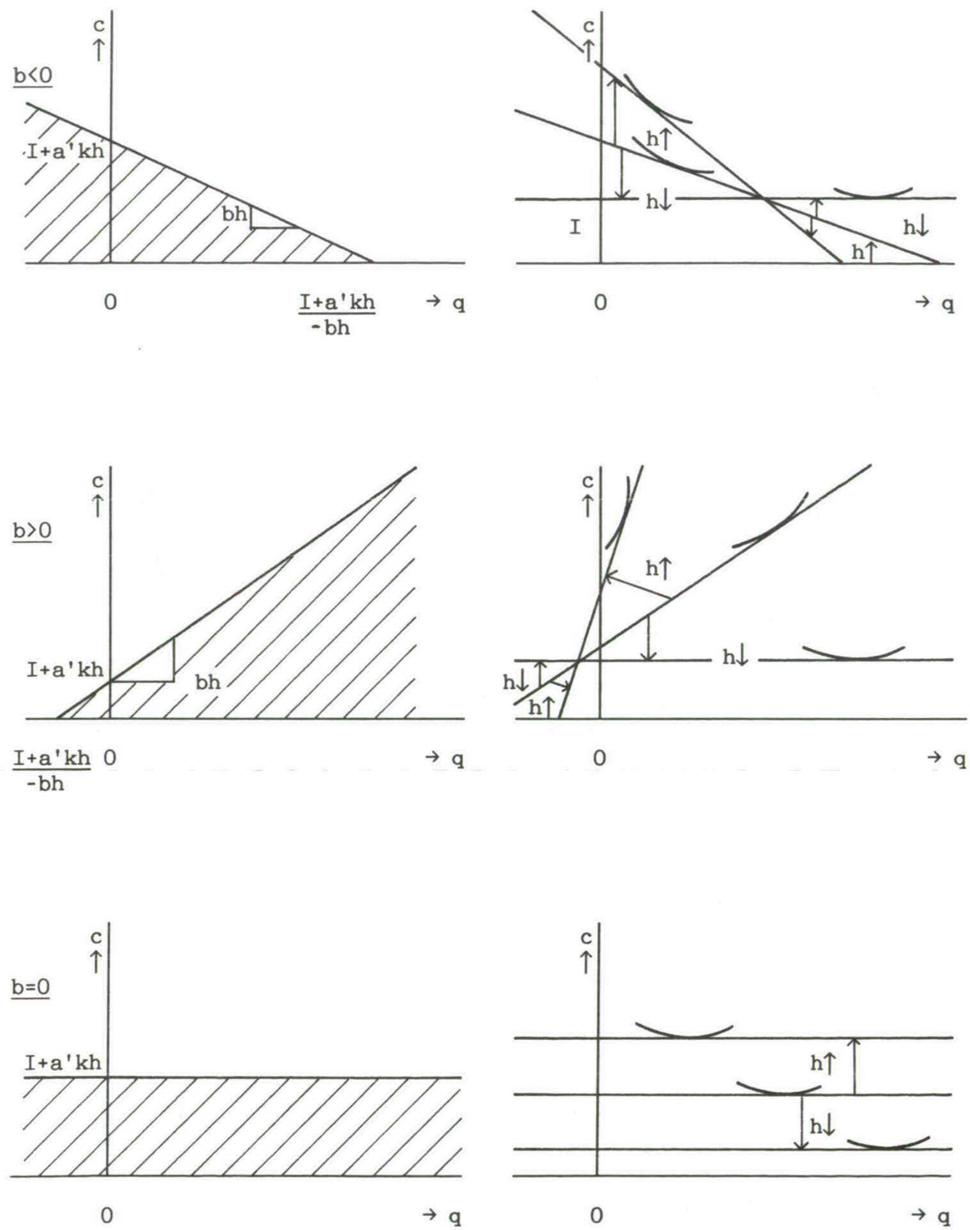


Figure 5.2 <u>The budget set in the (c,q)-plane for a given value of L</u> <u>The optimum situation in the (c,q)-plane for different values of L</u>

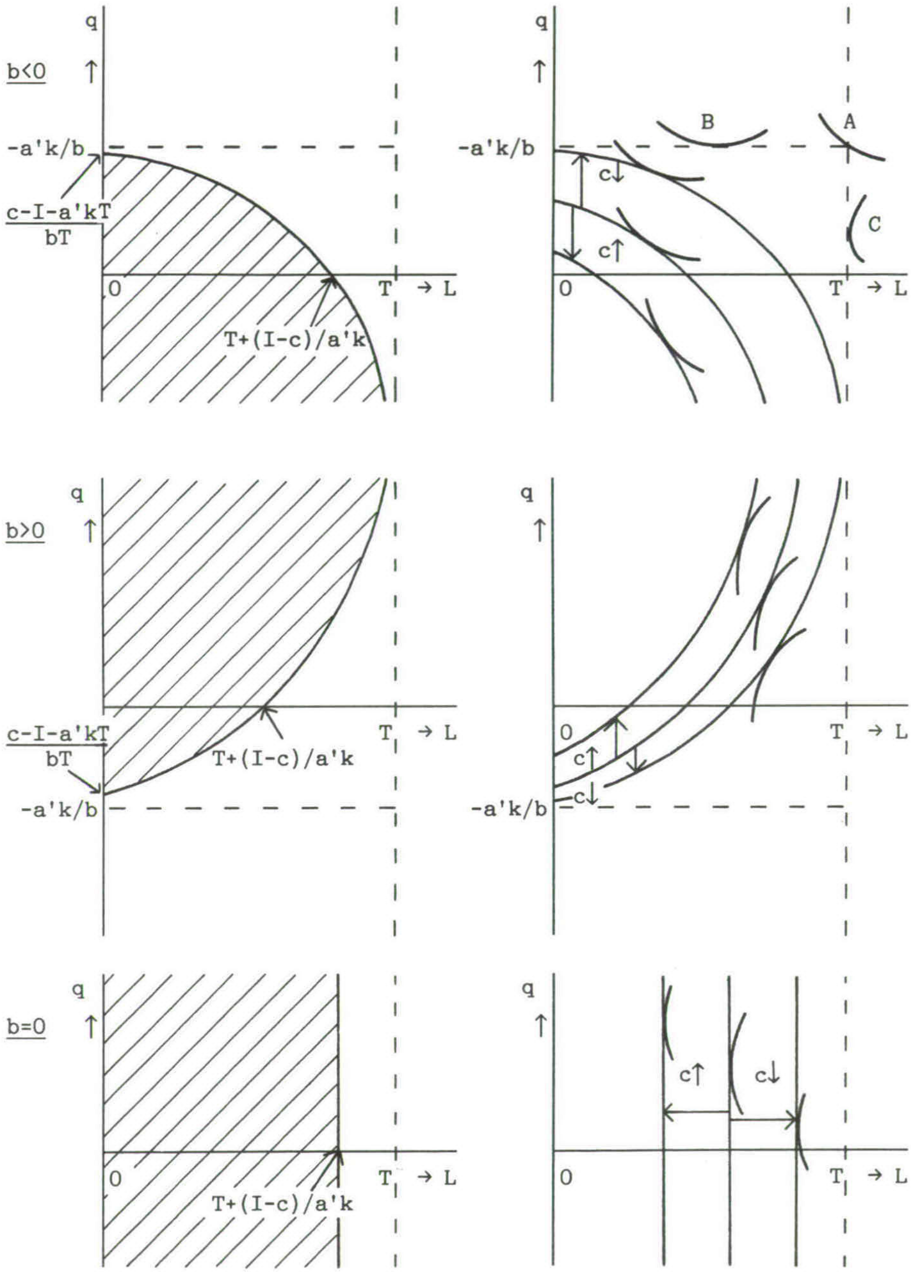


Figure 5.3 <u>The budget set in the (q,L)-plane for a given value of c</u>

<u>The optimum situation in the (q,L) plane for different values of c</u>

We now turn to the specification of the indirect utility function which is the same as the one used in the previous specification

$$V(\pi^*, IC^*) = \exp((\beta_1\pi_1^* + \beta_2\pi_2^*)/\pi_3^*)I^* \tag{5.56}$$

$$I^* = IC^*/\pi_3^* + \vartheta + \delta_1\pi_1^*/\pi_3^* + \delta_2\pi_2^*/\pi_3^* + 1/2\gamma_1(\pi_1^*/\pi_3^*)^2 + 1/2\gamma_2(\pi_2^*/\pi_3^*)^2 \tag{5.57}$$

Application of Roy's Identity and translating the system back into the goods-space yields demand functions for c,h and q

$$h = \beta_1\bar{I} + \delta_1 + \gamma_1 w \tag{5.58}$$

$$q = -\beta_2\bar{I} - \delta_2 - \gamma_2 p_q \tag{5.59}$$

$$c = (\beta_1 w + \beta_2 p_q + 1)\bar{I} - \vartheta + 1/2\gamma_1 w^2 + 1/2\gamma_2 p_q^2 \tag{5.60}$$

where $$\bar{I} = M + \vartheta + \delta_1 w + \delta_2 p_q + 1/2\gamma_1 w^2 + 1/2\gamma_2 p_q^2 \tag{5.61}$$

$$w = a'k + bq \tag{5.62}$$

$$p_q = -hb \tag{5.63}$$

$$M = I - hbq \tag{5.64}$$

δ_1 and δ_2 are parameterized as in equations (5.38)-(5.39).

This is a nonlinear system of simultaneous equations which could be estimated in principle . However, we have to deal with the fact that the q variable, describing the job characteristics, is a qualitative variable and that no information on q is available for nonworkers. Our strategy has been to obtain the reduced form of the system (see Appendix 5A).

Stochastic specification of the two models

We specify stochastic versions of the reduced form systems of both models by adding a disturbance term to each equation. Due to the adding-up restriction one of the equations may be dropped in the estimation. We drop the equation for consumption and obtain all the parameter estimates from the other equations.

Summarizing we specify the following model

$$\begin{aligned} h^o &= \hat{h} + \varepsilon_h \quad \text{if } \hat{h} + \varepsilon_h \geq 0 \\ &= 0 \quad \text{if } \hat{h} + \varepsilon_h < 0 \end{aligned} \tag{5.65}$$

$$\begin{aligned} w^o &= \hat{w} + \varepsilon_w \quad \text{if } \hat{h} + \varepsilon_h \geq 0 \\ &= 0 \quad \text{if } \hat{h} + \varepsilon_h < 0 \end{aligned} \tag{5.66}$$

$$q = \hat{q} + \varepsilon_q \tag{5.67}$$

$$\begin{aligned} q^o &= 1 \quad \text{if } q > 0 \text{ and } \hat{h} + \varepsilon_h \geq 0 \\ &\ 0 \quad \text{if } q \leq 0 \text{ and } \hat{h} + \varepsilon_h \geq 0 \end{aligned} \tag{5.68}$$

$$\begin{bmatrix} \varepsilon_h \\ \varepsilon_w \\ \varepsilon_q \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_h^2 & \rho_{hw}\sigma_h\sigma_w & \rho_{hq} \\ & \sigma_w^2 & \rho_{wq}\sigma_w \\ & & 1 \end{bmatrix} \right] \tag{5.69}$$

Carets (^) denote reduced form solutions (equations (5A.1)-(5A.3) or (5A.15) in Appendix 5A), q^o is the observed dichotomously measured job characteristic and w^o is the observed wage rate. The random terms can only represent stochastic measurement errors. They cannot represent stochastic preferences, since a normally distributed stochastic preference term would lead to a nonnormally and nonadditive random term in the reduced form equation. In Appendix 5B the likelihood function is presented.

The model, developed above, is referred to as the "extended" model. We will contrast the extended model with a "standard" model, describing the case where q is excluded from the direct utility function, i.e. q is no longer a choice variable. In that case the following demand functions for h and c can be derived (see Section 5.2 and Chapter 2)

$$c = (\beta_1 w + 1)I^* - \vartheta + 1/2\gamma_1 w^2 \tag{5.70}$$

$$h = \beta_1 I^* + \delta_1 + \gamma_1 w \tag{5.71}$$

$$w = a'k + bq \tag{5.72}$$

$$I^* = I + \vartheta + \delta_1 w + 1/2\gamma_1 w^2 \tag{5.73}$$

5.3 Empirical results

We have considered working conditions as the appropriate objectively measurable job characteristic. Working conditions are "bad" if the work is dirty, physically strenuous or smells unpleasantly. A problem is the measurement of job characteristics. In most studies a single, composite, measure of job characteristics is constructed from responses to a series of questions (cf. Atrostic (1982)). This approach assumes that those responses are commensurable. The job characteristic in our data set is measured as a dichotomous variable. This variable is used as an indicator for the latent variable working conditions. To be able to estimate a model in which endogenous variables are measured dichotomously, it is necessary to obtain the reduced form, because we need an explicit expression for the joint probability that the appropriate dependent variables are one or zero. In Appendix 5A the derivation of the reduced form of the model is given. As in the previous two chapters we have used preferred hours as the appropriate endogenous hours variable. One main difference with the previous chapters is that the wage equation is estimated simultaneous with the hours (and job characteristic) equation.

Only the OSA-sample contains information on job characteristics, therefore we cannot present estimates for the SEP-sample, unlike in the previous chapters. Sample information is given in Table 5.1. From this table it can be seen that about 62% of the males and about 75% of the females report to work in jobs with good working conditions. One would expect individuals working in jobs with good working conditions to work more hours per week than individuals working in jobs with bad working conditions. This is true for all groups, except for single males. All different types of individuals, except single females, earn more if they work in nice jobs than if they work in jobs with bad working conditions. This is probably due to the fact that workers in nice jobs are also the higher educated ones. One should note that the number of individuals working in jobs with bad working conditions is very small, especially for single males and females, and to a lesser extent for females in families. Therefore, we should be careful with the interpretation of the estimation results. Nevertheless, we present all estimation results in order to keep in line with the previous chapters of this thesis. But main emphasis will be put on the estimation results of the males in two-adult households.

The first columns of Tables 5.2 to 5.5 show results for an Ordinary Least Squares Regression of wages on a vector of human capital variables and on a dummy variable for good working conditions. The dummy equals one if the working conditions are good and zero otherwise, which means that the job characteristic is defined to be a "good". Following the usual estimated wage differential literature, this equation is not corrected for selection bias (contrary to the wage equation estimated in Chapter 2). The results of the maximum likelihood estimation of the standard and the extended models are presented in the other three columns of the tables. As was discussed in Section 2, job choice was assumed to be exogenous in the standard model. This model consists of two equations, an hours equation and a wage equation. In this way possible selection bias in the wage equation due to the selection of workers only, is taken care of. In the extended model job choice was made endogenous, thus extending the model with a job choice equation and hence deleting selection bias due to the self selection of individuals into different occupations.

Let us first turn to the estimation results of the males in families. All coefficients in the first two columns have the usual signs

- the higher nonlabour income is the more leisure is prefered ($\beta_1<0$)
- the higher the family size the more the male works ($\delta_{11}>0$)
- the linear wage effect on male hours is very small (γ_1 is on its lower bound 0.01)

- the higher one's education the higher one's wage rate ($a_{22}<a_{23}<a_{24}$)
- up to 50 years the higher one's age the higher one's wage rate, if one gets older than fifty the wage rate starts decreasing ($a_{31}>0$, $a_{32}<0$)
- a technical education doesn't increase one's wage ($a_{41}<0$) but an economic one does ($a_{42}>0$)
- Supervisors earn more than others ($a_5>0$) and so do individuals working in private firms, compared to individuals working in the civil service (a_6)

Turning to the interpretation of the results in the last two columns one should bear in mind that in the extended model the budget constraint is nonlinear and that the estimated parameters are just particular parameters of the utility function and of the budget constraint. For simplicity we will nevertheless mainly use conventional terminology to describe the parameters. With respect to the parameters in the hours equation and the wage equation the parameter estimates resemble those in the first two columns. As expected there is a positive income effect on the demand for the job characteristic, $\beta_2<0$, (this result is also found in Atrostic (1982)). Individuals with a high income can best afford non-pecuniary job characteristics. Age doesn't have a significant effect on the demand for the commodity Z_2 (i.e. the amenity of a job). But the higher one's education the higher the demand for Z_2 ($0<-\delta_{22}<-\delta_{23}<-\delta_{24}$). This holds for both extended models, although the coefficients are insignificant in the $Z_2=q$ version.

For the other groups roughly the same results hold. Of course for females the coefficient for family size is negative in the labour supply equation. On average the significant coefficients are relatively stable accross the different models, except the coefficient for education and the estimated wage differential. It is striking to see that for all different groups of individuals the effect of education on the demand for Z_2 is strongly positive and significant if $Z_2=hq$, and insignificant if $Z_2=q$. One explanation could be that it is not the job characteristic q but hours of work (and thus also hq) that depends on the level of education.

Both Atrostic (1982) and Killingsworth (1984) found that wage effects on male labour supply increased, when one went from a conventional labour supply model that ignored differences in job characteristics to a model in which such characteristics were treated as choice variables in a demand system. There is a slight increase in the wage effects in our models for males, and a large increase for married females in the Z_2=q version (see Table 5.6). This can also be concluded from the labour supply curves in the Figures 5.4, 5.6 and 5.8.

For married males the estimated compensating wage differential (b) in the OLS-regression and in the standard model is positive, but insignificant. In much empirical research on wage differentials this same phenomenon is found. According to the theory of compensating wage differentials, one would expect a negative sign for this coefficient. However it could be argued that the coefficient should not be interpreted as a compensating wage differential, but as a measure of the impact of omitted variables, such as ability and motivation. Whatever the interpretation, the two estimates of this coefficient are biased if one believes that the job characteristic is an endogenous variable. In the extended model with Z_2=hq the estimated wage differential is also insignificant. In the extended model with Z_2=q however, the estimated wage differential is negative and significant. Note that the interpretation of the wage differential is slightly different in the extended model from the one in the standard model, since b is the coefficient of the latent variable q and not of the observed variable q. All other coefficients in the first column have the usual sign. What are the implications of the estimates with respect to the reservation wage differential? The reservation wage differential is the amount of money that makes a representative individual (someone with sample mean values for all variables and zero values for all unobservables) indifferent between work with good working conditions and work with bad working conditions. The slope of the budget line in the optimum point is a usefull approximation. Married males are willing to accept a pay cut of 6.6% in order to take a job with better working conditions, according to the Z_2=q specification.

For single males and females the coefficient b is insignificant in all four models. This could be due to the small number of individuals working in different types of jobs. For females in families, the estimated wage differential is positive and significant implying that females working in jobs with good working conditions are better paid than those working in jobs with bad working conditions. The positive value of b can be more reasonably interpreted as the measurement of the impact of omitted variables, such as motivation.

With respect to the simulated male hours distribution there is hardly any difference between the three models (compare Figures 5.5, 5.7 and 5.9). All models underpredict the degree of nonparticipation. For single females the same phenomenon can be seen. For married females the first two models simulate the number of nonparticipants better than the models described in earlier chapters. The Z_2=q version, however, largely overpredicts the number of nonparticipants. Adding the random term spreads out the hours distribution, just as we have seen in the previous chapters.

Table 5.1 Means of some main variables

males

	single			families		
	total	q=0	q=1	total	q=0	q=1
q	0.71	0	1	0.60	0	1
h^p	35.8	38.3	34.8	38.7	38.4	38.9
w	15.7	13.5	16.5	16.0	14.7	16.8
N	111	32	79	799	316	483

females

	single			families		
	total	q=0	q=1	total	q=0	q=1
q	0.79	0	1	0.74	0	1
h^p	32.8	30.7	33.4	24.5	24.4	24.6
w	13.5	14.3	13.3	12.5	11.6	12.8
N	127	27	100	318	84	234

where q = 1 if the working conditions are good
= 0 if the working conditions are bad
N = number of working individuals

Table 5.2 Estimation results for single males[a)]

	OLS Regression	Standard Model	Extended Model Z_2=hq	Extended Model Z_2=q
ϑ		-1650(fixed)	-1650(fixed)	-1650(fixed)
δ_{10}		35.0(0.8)	35.0(0.9)	32.5(4.2)
β_1		-0.0002(0.0002)	-0.0001(0.0002)	-0.000001(ℓ.b.)
δ_{11}		3.7(3.5)	3.8(3.7)	0.6(4.1)
γ_1		0.01(ℓ.b.)	0.01(ℓ.b.)	0.2(0.3)
σ_h		5.9(0.4)	6.0(0.4)	6.0(0.3)
β_2			-0.0002(0.0005)	-0.00003(0.00004)
δ_{21}			-0.8(0.8)	0.002(0.02)
δ_{22}			-21.1(26.6)	0.4(0.4)
δ_{23}			-47.6(23.7)	0.7(0.7)
δ_{24}				
γ_2			1.1(2.0)	0.01(ℓ.b.)
a_1	-10.9(12.3)	-11.1(21.5)	-66.6(74.8)	-9.4(19.9)
a_{22}	-2.7(3.0)	-2.7(5.3)	22.2(44.1)	-3.0(6.4)
a_{23}	1.2(2.7)	1.2(3.5)	57.8(72.1)	-4.9(3.8)
a_{24}				
a_{31}	1.1(0.6)	1.1(1.1)	2.2(1.8)	1.3(1.0)
a_{32}	-0.01(0.01)	-0.01(0.01)	-0.01(0.02)	-0.01(0.02)
a_{41}	1.0(1.6)	1.0(2.8)	1.0(2.6)	1.5(2.1)
a_{42}	4.7(1.8)	4.6(2.4)	4.4(2.4)	5.2(1.9)
a_5				
a_6	0.4(1.4)	0.4(1.7)	0.8(1.8)	0.8(1.8)
b	1.6(1.6)	1.5(2.8)	-42.5(57.1)	-0.5(3.7)
σ_w	6.9(0.2)	6.6(0.3)	6.5(0.3)	6.0(0.3)
ρ_{hw}		-0.01(0.18)	-0.0(0.2)	-0.1(0.2)
ρ_{hq}			-0.3(0.1)	-0.4(0.1)
ρ_{wq}			0.3(0.3)	0.2(0.2)
log lik	-360.5	-507.4	-563.2	-565.6

a) Standard errors in parentheses

Table 5.3 Estimation results for single females[a)]

	OLS Regression	Standard Model	Extended Model Z_2=hq	Extended Model Z_2=q
ϑ		-355(fixed)	-355(fixed)	-355(fixed)
δ_{10}		18.7(10.2)	34.2(6.1)	18.3(10.3)
β_1		-0.0003(0.0001)	-0.0004(0.0001)	-0.000001(u.b.)
δ_{11}		-5.6(2.8)	-3.5(2.7)	-9.4(2.9)
γ_1		1.2(0.8)	0.04(0.2)	1.2(0.7)
σ_h		7.5(0.7)	7.7(0.7)	7.7(0.7)
β_2			0.0006(0.0003)	-0.0001(0.00002)
δ_{21}			-0.3(0.2)	0.01(0.01)
δ_{22}			-11.1(17.4)	0.2(0.5)
δ_{23}			-22.2(9.0)	-0.4(0.5)
δ_{24}				
γ_2			0.01(ℓ.b.)	0.01(ℓ.b.)
a_1	6.8(7.2)	5.3(11.0)	0.7(12.0)	3.4(8.0)
a_{22}	1.3(2.3)	0.4(4.0)	7.4(10.4)	0.5(1.5)
a_{23}	1.9(1.8)	0.6(3.5)	11.6(13.5)	-1.8(3.7)
a_{24}				
a_{31}	0.2(0.4)	0.4(0.5)	0.6(0.5)	0.7(0.4)
a_{32}	-0.002(0.01)	-0.005(0.006)	-0.005(0.006)	-0.009(0.005)
a_{41}	2.2(2.9)	1.1(3.9)	2.1(3.8)	2.3(2.5)
a_{42}	-0.5(1.4)	-0.8(1.9)	-0.5(1.9)	-0.2(1.1)
a_5				
a_6	0.4(1.2)	0.3(1.9)	0.6(1.8)	0.02(1.1)
b	-1.1(1.3)	-0.4(1.9)	-13.3(15.9)	-1.3(1.4)
σ_w	6.1(0.2)	5.9(0.3)	5.8(0.3)	6.1(0.3)
ρ_{hw}		-0.24(0.13)	-0.2(0.2)	-0.3(0.1)
ρ_{hq}			0.2(0.2)	0.2(0.2)
ρ_{wq}			-0.1(0.2)	-0.1(0.2)
log lik	-345.9	-588.5	-651.3	-657.2

a) Standard errors in parentheses

Table 5.4 Estimation results for males in families[a)]

	OLS Regression	Standard Model	Extended Model Z_2=hq	Extended Model Z_2=q
ϑ		-1650(fixed)	-1650(fixed)	-1650(fixed)
δ_{10}		34.7(2.0)	34.7(2.0)	43.9(5.8)
β_1		-0.002(0.001)	-0.001(0.001)	-0.0001(0.0001)
δ_{11}		2.0(0.7)	2.1(0.7)	-0.3(0.5)
γ_1		0.01(ℓ.b.)	0.01(ℓ.b.)	0.2(0.3)
σ_h		6.0(0.1)	6.0(0.1)	6.0(0.1)
β_2			-0.02(0.01)	-0.0004(0.0003)
δ_{21}			-0.1(0.1)	0.001(0.002)
δ_{22}			-21.4(6.0)	0.1(0.2)
δ_{23}			-30.6(5.2)	-0.2(0.3)
δ_{24}			-56.4(6.0)	-0.3(0.4)
γ_2			0.01(ℓ.b.)	0.01(ℓ.b.)
a_1	-1.9(3.0)	-2.0(3.6)	-0.5(4.2)	-1.6(5.0)
a_{22}	0.6(0.6)	0.6(0.8)	0.2(1.3)	4.3(2.2)
a_{23}	1.6(0.6)	1.7(0.7)	1.0(1.7)	7.0(2.8)
a_{24}	4.9(0.6)	4.8(0.7)	3.5(3.0)	15.5(5.3)
a_{31}	0.5(0.1)	0.5(0.2)	0.5(0.2)	0.4(0.2)
a_{32}	-0.005(0.002)	-0.005(0.002)	-0.005(0.002)	-0.003(0.003)
a_{41}	-0.1(0.4)	-0.1(0.4)	-0.2(0.4)	-0.3(0.5)
a_{42}	1.7(0.5)	1.7(0.5)	1.9(0.5)	2.3(0.7)
a_5	1.1(0.3)	1.2(0.3)	1.2(0.3)	1.3(0.3)
a_6	1.0(0.4)	1.1(0.4)	0.9(0.4)	1.0(0.4)
b	0.2(0.4)	0.2(0.5)	0.8(1.8)	-6.6(3.2)
σ_w	5.0(0.02)	5.0(0.1)	5.0(0.1)	5.0(0.1)
ρ_{hw}		-0.1(0.04)	-0.1(0.04)	-0.1(0.04)
ρ_{hq}			0.1(0.05)	0.1(0.05)
ρ_{wq}			0.01(0.05)	0.02(0.05)
log lik	-2150.5	-3396.6	-3865.8	-3861.6

a) Standard errors in parentheses

Table 5.5 Estimation results for females in families[a)]

	OLS Regression	Standard Model	Extended Model Z_2=hq	Extended Model Z_2=q
ϑ		-355(fixed)	-355(fixed)	-355(fixed)
δ_{10}		0.0(fixed)	0.0(fixed)	0.0(fixed)
β_1		-0.016(0.003)	-0.000001(u.b.)	-0.000001(u.b.)
δ_{11}		-25.4(2.8)	-11.3(1.2)	-13.5(3.1)
γ_1		3.4(0.2)	4.1(0.3)	3.1(0.3)
σ_h		10.5(0.5)	10.2(0.5)	8.7(0.5)
β_2			-0.001(0.003)	0.00003(0.00001)
δ_{21}			0.10(0.06)	-0.02(0.002)
δ_{22}			-3.0(1.5)	0.1(0.1)
δ_{23}			-5.5(1.4)	0.2(0.1)
δ_{24}			-15.8(2.9)	0.5(0.1)
γ_2			0.9(0.4)	0.001(0.001)
a_1	-6.4(3.2)	-5.3(2.5)	-1.8(1.6)	-2.9(5.6)
a_{22}	1.0(0.7)	0.0(0.5)	0.3(0.2)	-0.3(1.4)
a_{23}	1.5(0.6)	1.2(0.4)	1.1(0.2)	1.7(1.4)
a_{24}	4.7(0.8)	3.6(0.5)	2.6(0.4)	5.7(2.1)
a_{31}	0.9(0.2)	0.5(0.1)	0.2(0.1)	-0.1(0.3)
a_{32}	-0.01(0.002)	-0.01(0.002)	-0.003(0.001)	-0.003(0.003)
a_{41}	-1.5(1.4)	0.1(1.3)	-0.1(0.6)	0.3(4.7)
a_{42}	0.8(0.7)	-0.3(0.6)	-0.0(0.5)	-1.1(0.6)
a_5	0.0(0.5)	1.2(0.4)	0.9(0.3)	1.9(1.7)
a_6	-0.5(0.5)	4.0(0.4)	4.2(0.4)	7.5(1.8)
b	0.8(0.5)	2.5(0.4)	8.1(1.3)	11.6(3.6)
σ_w	3.8(0.2)	4.6(0.2)	4.2(0.1)	4.5(0.1)
ρ_{hw}		0.2(0.1)	-0.001(0.01)	-0.06(0.06)
ρ_{hq}			-0.1(0.1)	0.1(0.1)
ρ_{wq}			0.2(0.1)	0.1(0.1)
log lik	-770.6	-1157.8	-1349.1	-1215.7

a) Standard errors in parentheses

Table 5.6 Wage elasticities[a)]

	standard		extended (Z_2=hq)		extended (Z_2=q)	
	singles	families	singles	families	singles	families
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.00	-0.02	0.00	-0.02	0.00	-0.00
	(0.00)	(0.02)	(0.00)	(0.03)	(0.13)	(0.00)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.51	1.64	0.01	1.64	0.51	2.37
	(0.32)	(0.34)	(0.11)	(0.08)	(0.34)	(1.59)

a) Standard errors in parentheses

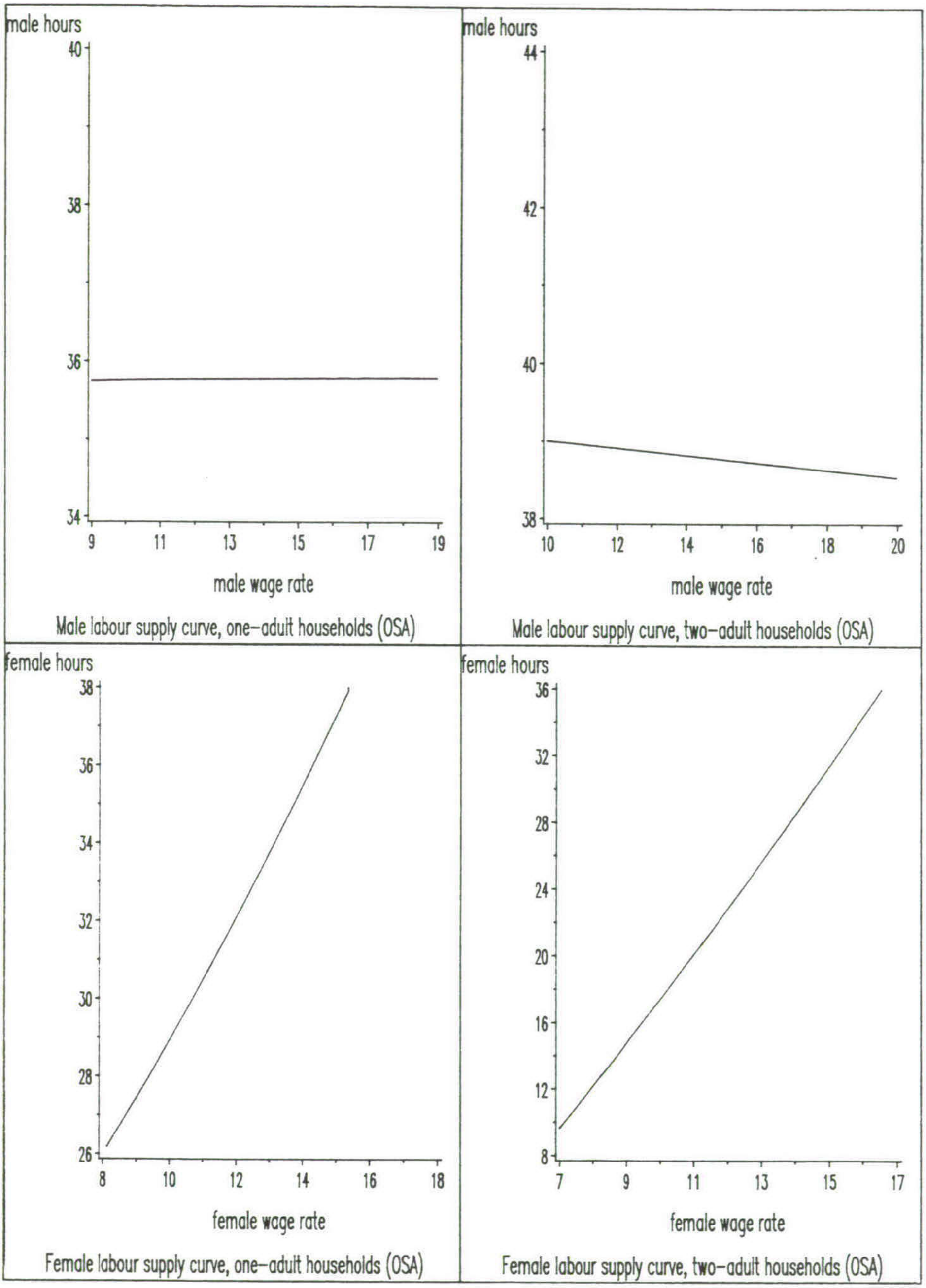


Figure 5.4 <u>Labour supply curves, standard model (OSA)</u>

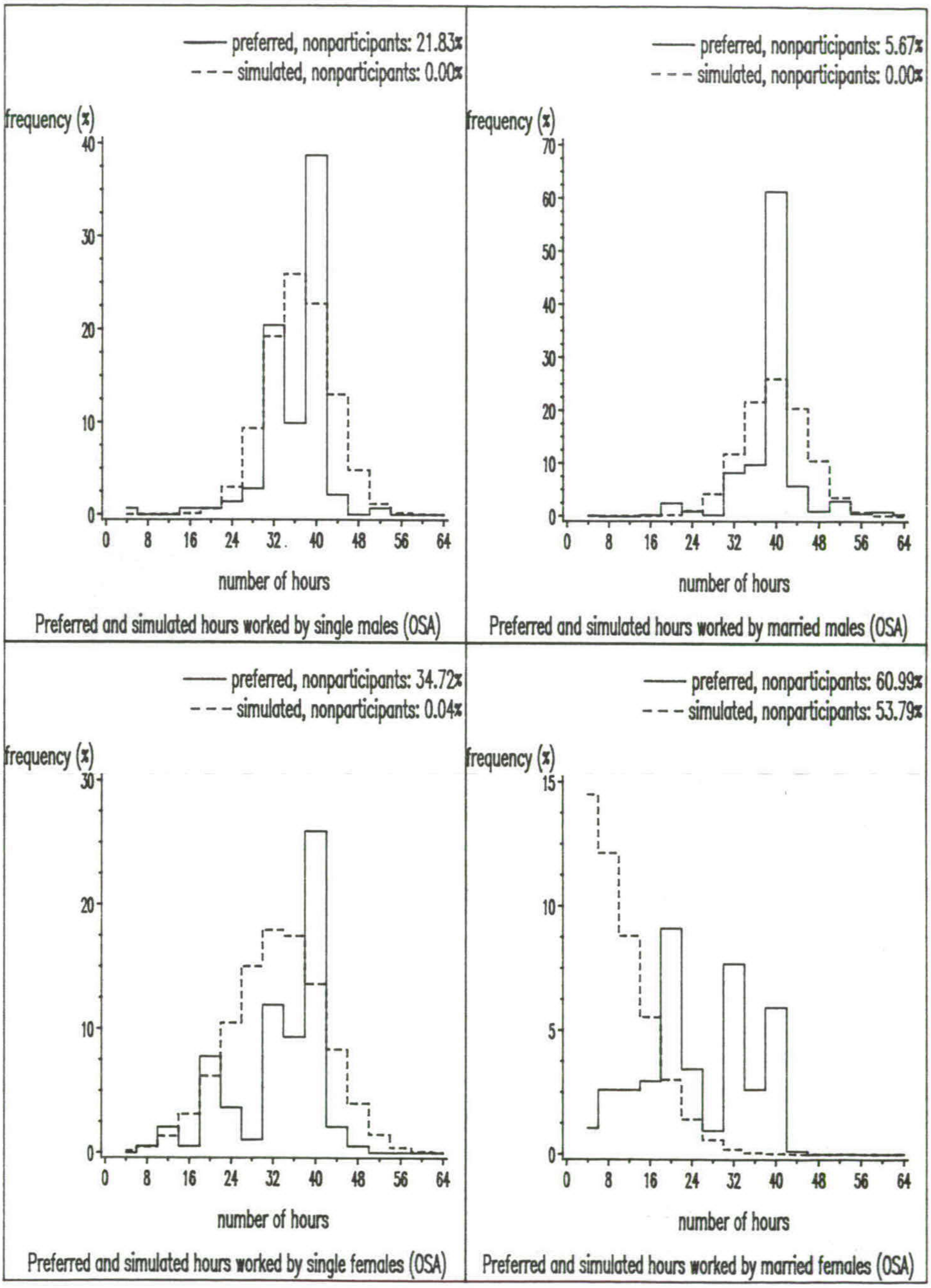


Figure 5.5 <u>Hours distributions, standard model (OSA)</u>

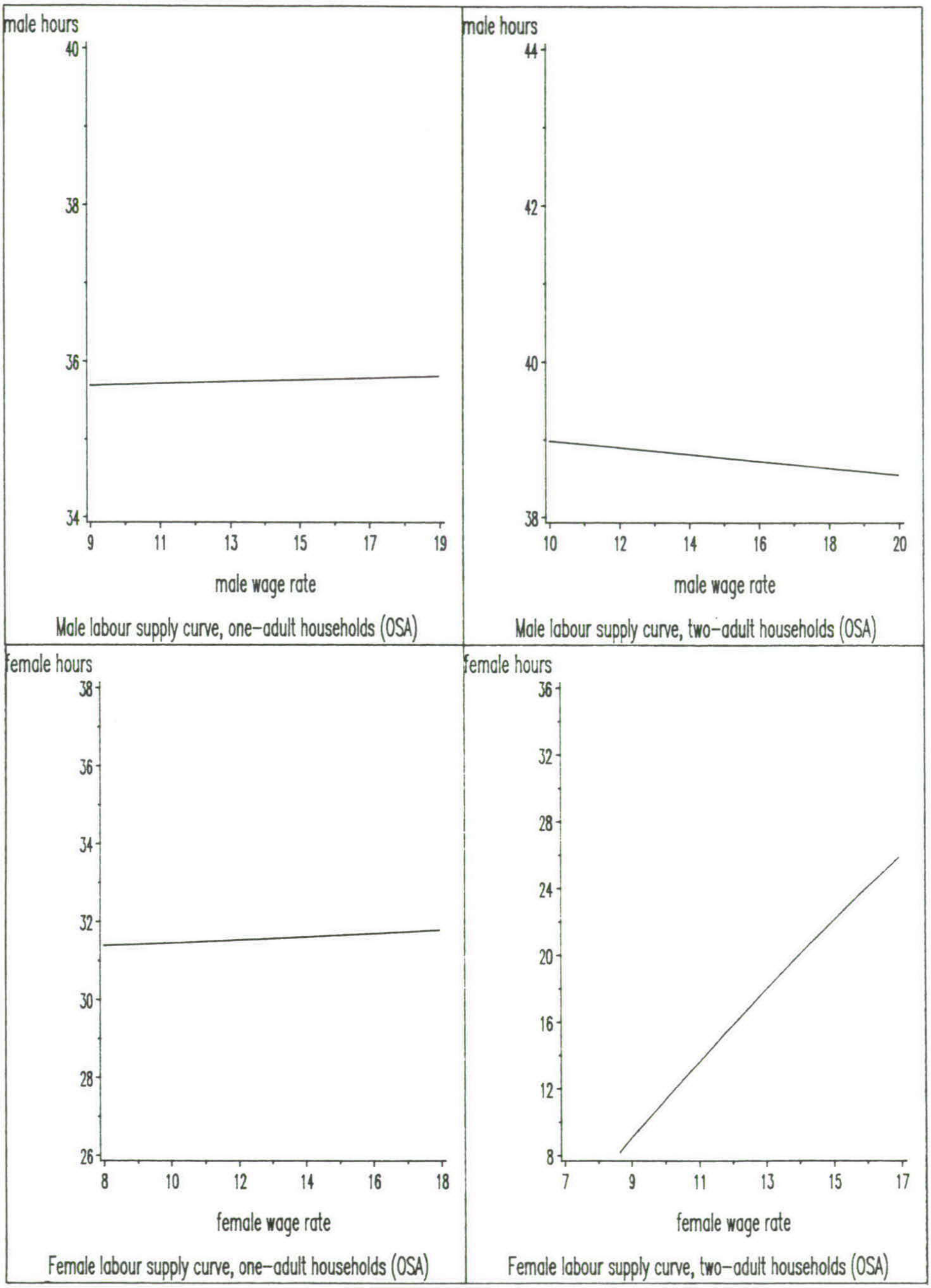


Figure 5.6 Labour supply curves, Z_2=hq (OSA)

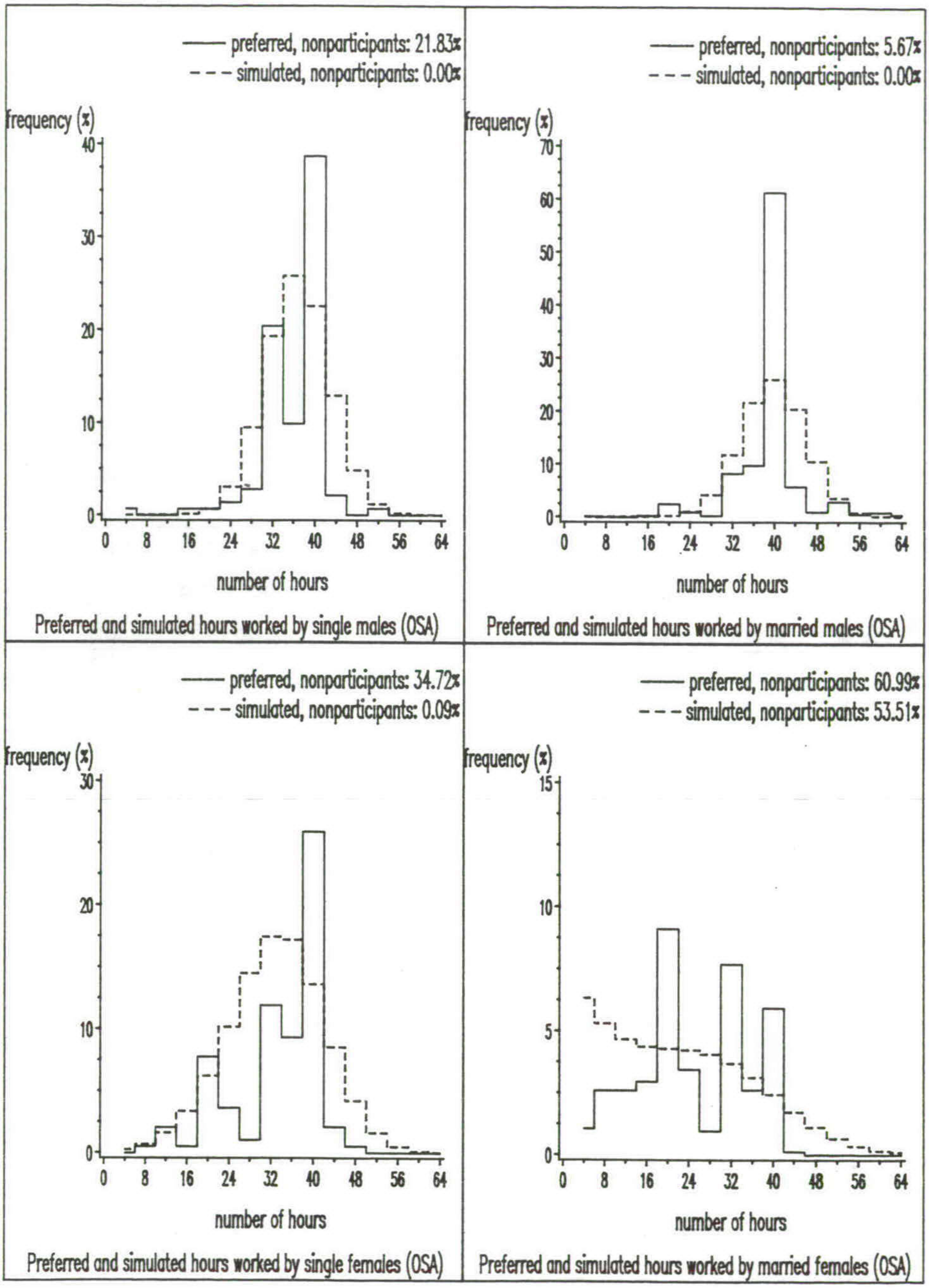


Figure 5.7 Hours distributions, Z_2=hq (OSA)

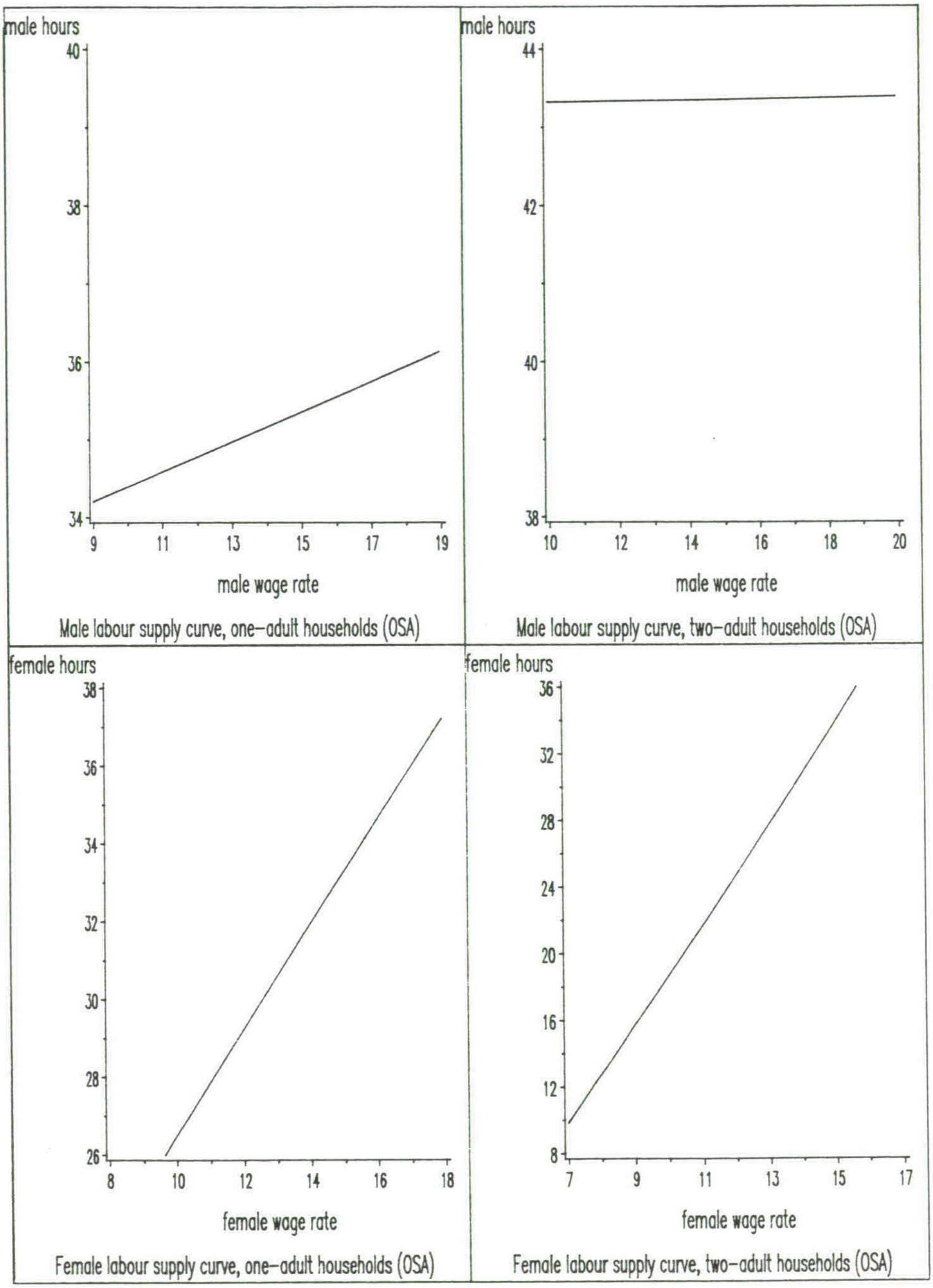


Figure 5.8 Labour supply curves, Z_2=q (OSA)

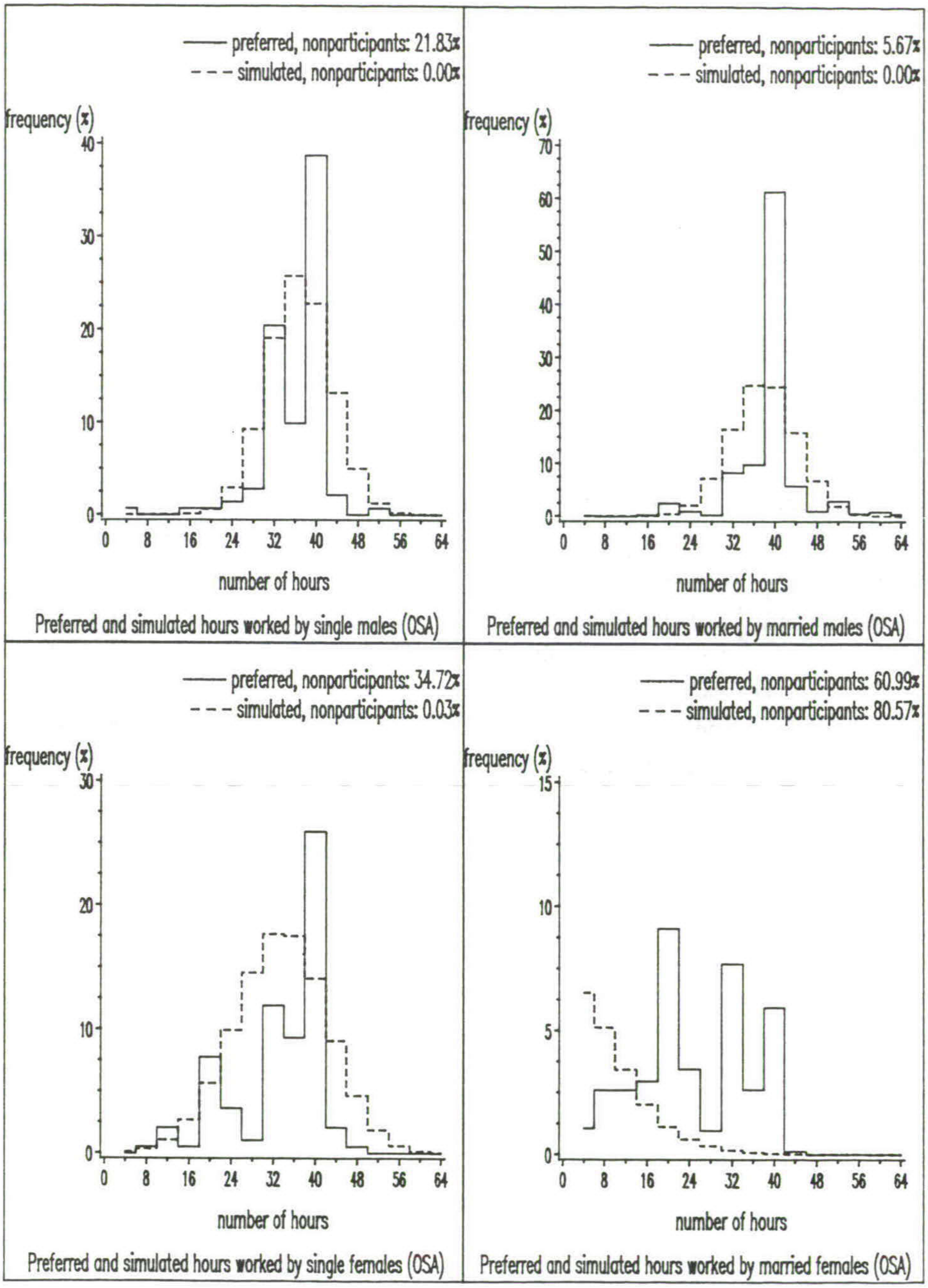


Figure 5.9 <u>Hours distributions, Z_2=q (OSA)</u>

5.4 Conclusion

In this chapter a structural model of job choice, labour supply and wages has been developed. The household production approach provides the theoretical background. An individual is assumed to maximize a utility function with leisure, a nonpecuniary job characteristic and consumption as arguments. Moreover it is assumed that one's wage depends on human capital variables and the characteristics of the job actually chosen. As a result the wage rate is endogenous and the budget constraint is nonlinear.

The estimation results are not uniformly satisfactory. Except for the married males the number of individuals working in jobs with bad working conditions is small. No strong conclusions can be drawn for these groups. According to the extended model with specification Z_2=q the reservation wage differential is 6.6 for married males. This implies that the individual is prepared to loose 6.6% of his real income to obtain a job with better working conditions. This model yields estimation results that differ in sign and in magnitude from results obtained for models that do not take into account the endogeneity of job choice. Therefore, analyses of labour supply and wages that ignore the (self) selection of individuals into different occupations, may produce misleading results.

Appendix 5A Derivation of the reduced form

Specification Z_2=hq

Starting from equations (5.33)-(5.36) the reduced form of this model is very easily derived by dividing the expression of hq by the expression of h. We then obtain the reduced form model we have estimated

$$\hat{h} = \beta_1\bar{I} + \delta_1 + \gamma_1 a'k \tag{5A.1}$$

$$\hat{q} = (-\beta_2\bar{I} - \delta_2 + \gamma_2 b)/(\beta_1\bar{I} + \delta_1 + \gamma_1 a'k) \tag{5A.2}$$

$$\hat{w} = a'k + b(-\beta_2\bar{I} - \delta_2 + \gamma_2 b)/(\beta_1\bar{I} + \delta_1 + \gamma_1 a'k) \tag{5A.3}$$

where $$\bar{I} = I+\vartheta+\delta_1 a'k-\delta_2 b+1/2\gamma_1(a'k)^2+1/2\gamma_2 b^2 \tag{5A.4}$$

Specification Z_2= q

We start from the demand equations in the goods space (5.58)-(5.64). First we shall derive a reduced form quadratic equation in h only. Given a solution for h, (5.59) immediately yields a solution for q, and thus for w. From (5.59) it follows that

$$bq = -b\beta_2\bar{I} - b\delta_2 + hb^2\gamma_2 \tag{5A.5}$$

Rewriting (5.58) yields

$$\frac{(h - \delta_1 - \gamma_1(a'k + bq))}{\beta_1} = \bar{I} \tag{5A.6}$$

Substitution of equation (5A.6) into (5A.5) gives

$$bq = c_1 h + c_2 \tag{5A.7}$$

where $$c_1 = \frac{b\ \beta_2 - \beta_1 b^2\gamma_2}{(\gamma_1 b\ \beta_2 - \beta_1)} \tag{5A.8}$$

$$c_2 = \frac{\beta_1 b\ \delta_2 - b\beta_2(\delta_1 + \gamma_1 a'k)}{(\gamma_1 b\ \beta_2 - \beta_1)} \tag{5A.9}$$

Substitution of equation (5A.7) into (5.58) yields a quadratic equation in h

$$d_1 h^2 + d_2 h + d_3 = 0 \tag{5A.10}$$

where $d_1 = \beta_1(-c_1 + 1/2\gamma_1 c_1^2 + 1/2\gamma_2 b^2)$ (5A.11)

$$d_2 = -1 + \beta_1(-c_2 + \delta_1 c_1 - \delta_2 b + \gamma_1 a'kc_1 + \gamma_1 c_1 c_2) + \gamma_1 c_1 \tag{5A.12}$$

$$d_3 = \beta_1(I + \vartheta + \delta_1 a'k + \delta_1 c_2 + 1/2\gamma_1(a'k)^2 + \gamma_1 a'kc_2 + 1/2\gamma_1 c_2^2) + \delta_1 + \gamma_1 a'k + \gamma_1 c_2 \tag{5A.13}$$

The two solutions of equation (5A.10) are

$$h = \frac{-d_2 \pm \sqrt{D}}{2d_1} \tag{5A.14}$$

where $D = d_2^2 - 4d_1 d_3$ (5A.15)

Substitution of (5A.14) into (5.59) yields two solutions for the vector q, corresponding with the two solutions for h, given in (5A.14). Equation (5A.10) has a real solution if and only if the discriminant $D \geq 0$. This restriction is imposed in estimation.

Some theoretical implications of utility maximization with a nonlinear budget constraint

One condition that should hold for both specifications is that the budget curve is more convex than the indifference curve. Convexity of the indifference curve is equivalent with concavity of the cost function. The cost function corresponding to the indirect utility function given in (5.28)-(5.29) or in (5.56)-(5.57) is

$$C(\pi) = u.\exp(-\pi'p) - \vartheta - \delta'\pi - 1/2\pi'A\pi \qquad (5A.16)$$

where u denotes the individual's utility level and π is the price vector. Concavity of the cost function requires that

$$E = (I + \vartheta + \delta'\pi + 1/2\pi'A\pi)\beta\beta' - A \qquad (5A.17)$$

is negative semi-definite. Assuming that A is positive definite, a necessary condition for concavity is

$$(I + \vartheta + \delta'\pi + 1/2\pi'A\pi)\beta'A^{-1}\beta - 1 \leq 0 \qquad (5A.18)$$

Condition (5A.18), rewritten in terms of h, is checked in estimation.

Appendix 5B The likelihood function

Standard model

The likelihood contribution of a working individual is:

$$L^1 = \frac{1}{\sigma_h \sigma_w} b\varphi \left[\frac{h^o - \hat{h}}{\sigma_h}, \frac{w^o - \hat{w}}{\sigma_w}, \rho_{hw} \right] \tag{5B.1}$$

The likelihood contribution of a nonworking, seeking individual is

$$L^2 = \Phi \left[\frac{\hat{h}}{\sigma_h} \right] \tag{5B.2}$$

The likelihood contribution of a nonworking, nonseeking individual is

$$L^3 = \Phi \left[\frac{-\hat{h}}{\sigma_h} \right] \tag{5B.3}$$

Extended model

The likelihood contribution for working individuals with $q^o = 1$ is

$$f_1(h,w) \int_0^{\infty} f_c(q|\ h,w)dq \tag{5B.4}$$

where f_1 is the marginal density of h and w
f_c is the conditional density of q, given h and w

and for working individuals with $q^o = 0$

$$f_1(h,w) \int_{-\infty}^{0} f(q|\ h,w)dq \tag{5B.5}$$

Equation (5B.4) can be written explicitly as

$$L^3 = L^1 \, \Phi \left[\frac{\mu_q}{\sigma_q} \right] \tag{5B.6}$$

where $$\mu_q = \hat{q} + \left[\left[\frac{h^o - \hat{h}}{\sigma_h} \right] (\rho_{hq} - \rho_{hw}\rho_{wq}) + \left[\frac{w^o - \hat{w}}{\sigma_w} \right] (\rho_{wq} - \rho_{hw}\rho_{hq}) \right] / (1 - \rho_{hw}^2) \tag{5B.7}$$

$$\sigma_q^2 = 1 - \left(\rho_{hq}^2 + \rho_{wq}^2 - 2\rho_{hw}\rho_{hq}\rho_{wq} \right) / (1 - \rho_{hw}^2) \tag{5B.8}$$

Likewise we can write equation (5B.5) as

$$L^4 = L^1 \, \Phi \left[\frac{-\mu_q}{\sigma_q} \right] \tag{5B.9}$$

The likelihood contribution of a nonworking, seeking individual is

$$L^5 = \Phi \left[\frac{\hat{h}}{\sigma_h} \right] \tag{5B.10}$$

The likelihood contribution of a nonworking, nonseeking individual is

$$L^6 = \Phi \left[\frac{-\hat{h}}{\sigma_h} \right] \tag{5B.11}$$

6 Hours restrictions

6.1 Introduction

In the preceding chapters various models of labour supply were presented. In this chapter we will in addition model demand side restrictions, although in a simple way. Our starting point is an individual (no household) utility function that leads to a linear labour supply equation. Neither preference formation nor job characteristics nor the social security and welfare system are taken into account.

As we have seen in the preceding chapters the simulated distribution of hours in general fits the actual data on hours poorly. One reason for the bad fit of the hours distribution could be the invalid assumption of a fixed wage rate. Moffitt (1984), among others, extended the standard neoclassical labour supply model by making the wage rate endogenous and found significant effects of hours of work on the wage rate, leading to an S-shaped budget constraint. This was in support of the hypothesis put forward by Barzel (1973), namely that the marginal productivity (and thus the marginal wage rate) eventually declines at higher number of working hours. Rosen (1976) argued that the wage rate might depend on the number of working hours, due to the fact that there exist different markets for jobs with varying numbers of hours. And there is no reason that the market for full time jobs will clear at the same wage rate as the market for part time jobs. Especially in The Netherlands, where there is a growing interest in part time jobs, mainly by women, this might be an important consideration. Another reason for making the net wage rate dependent on hours of work is the progressive tax system.

Although the model with hours dependent wages fitted the hours distribution better than the standard Tobit model (Moffitt (1984), Tummers and Woittiez (1989)), the assumption of fixed wage rates does not seem to be the only invalid assumption. More important in this respect is the assumption that individuals can freely choose the number of hours they prefer to work. If the diversity of the offered hour packages is large enough, if workers have complete information about job opportunities, and if they are mobile between jobs, they will choose the job with exactly the number of hours they prefer. If workers are not perfectly mobile, for example, they may not be able to work their preferred number of hours. From the available job opportunities they will choose the number of hours yielding highest utility. To our knowledge the first study to estimate a model with hours restrictions on micro data is by Moffitt (1982). The way we have incorporated hours restrictions is largely based on an article by Dickens and Lundberg (1985). They present a model in which individuals may face constraints on their number of working hours. Their model is set up as a discrete choice model, in which each worker can choose from a limited number of job offers, with fixed numbers of hours. However, they assumed the wage rate to be fixed. In this chapter we build a model which incorporates both hours restrictions and hours dependent wages. By taking into account the availability of jobs with different numbers of hours and hours dependent wages we take a first step in the direction of modelling both the supply side and the demand side of the labour market.

In Section 6.2 we present a model with both hours restrictions and hours dependent wages. Hours restrictions are incorporated by assuming that employers offer jobs with a fixed number of hours. Workers face the market distribution of these employment opportunities. An individual chooses the number of hours corresponding with that one among the available job offers that yields highest utility. Notice that the individual is still a utility maximising person, although he maximises utility on a *subset* of all possible numbers of hours. This subset can be empty, because the number of job offers is a random variable of which zero is one of the possible outcomes. Consequently, the model distinguishes between voluntary and involuntary unemployment. Wages are made endogenous by specifying a wage equation in which the wage rate depends on hours of work and squared hours of work. In Section 6.3 estimation results will be presented and we discuss the improvement of fit of the model extension. Section 6.4 concludes.

6.2 The model

In this section we will first point out in what way hours restrictions can be incorporated in a standard labour supply model and then we discuss the implications of dropping the assumption that wages are independent of hours of work. For notational ease subscripts referring to individuals are omitted.

- Incorporation of hours restrictions

In contrast with the previous chapters we use a labour supply equation that is linear in wages. The main advantage of the linear specification is its computational simplicity. The obvious disadvantage is that the specification is less flexible. Starting point of the analysis is the following direct utility function (see Hausman (1980), Moffitt (1984))

$$\log(U(h,y)) = -\log(\gamma-\beta h) - \frac{\beta(h-X\delta-\varepsilon_h-\beta y)}{(\gamma-\beta h)} \tag{6.1}$$

where

h = working hours

y = disposable income

X = vector of individual characteristics such as age and family size

ε_h = random variable, representing unobserved tastes for work

γ, β and δ are parameters

$\gamma>0$, $\beta<0$

The restrictions $\gamma>0$ and $\beta<0$ are sufficient conditions for quasi-concavity of the utility function.

Maximizing the utility function (6.1) subject to a linear budget constraint yields a linear labour supply function

$$h = \beta I + \gamma w + X\delta + \varepsilon_h \tag{6.2}$$

where

I = nonlabour income

w = wage rate

The error term ε_h is a parameter of the utility function and hence is assumed to represent unobserved tastes for work. It cannot represent measurement errors. The wage equation is specified as

$$w = Z\psi + bh + ch^2 + \varepsilon_w \tag{6.3}$$

where

Z = vector of individual characteristics relevant for one's productivity, such as age and education

ε_w = error term

ψ, b, and c are parameters

The error term ε_w can be interpreted as a measurement error or as unobserved wage determinants such as ability.

We assume

$$\begin{bmatrix} \varepsilon_h \\ \varepsilon_w \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_h^2 & \rho_{hw}\sigma_h\sigma_w \\ \rho_{hw}\sigma_h\sigma_w & \sigma_w^2 \end{bmatrix}\right] \tag{6.4}$$

In estimation net wages are used, since individual labour supply is based on net wages rather than gross wages. By using net wages we implicitly estimate the tax system in the wage equation. In other words the wage-hours curvature may be due to tax-related nonlinearities. On the other hand gross wages should be used in estimating the demand side wage equation and then the hours-wage curvature is only due to non-tax related things. In modelling labour supply adjusting for both income taxes and hours restrictions, we refer to Van Soest, Woittiez, Kapteyn (1989).

If b and c are equal to zero, the wage rate is fixed and we are dealing with a linear budget constraint. Then the labour supply function given in (6.2) follows from maximizing the utility function (6.1). When the wage equation (6.3) is substituted into the budget constraint

$$y = wh + I \tag{6.5}$$

a nonlinear budget constraint results

$$y = hZ\psi + bh^2 + ch^3 + I + h\varepsilon_w \tag{6.6}$$

Maximizing the utility function (6.1) subject to the budget constraint (6.6) yields a nonlinear first order condition in the form of a third order polynomial in h. Estimation of this model would require analytical or numerical solutions to this cubic equation in h. But as will be explained presently we reformulate the model as a discrete choice problem in which utility is compared between a finite number of points of the budget constraint (0, h_1, h_2,..., h_m). Therefore it suffices to know the exact specification of the utility function.

We suppose an individual is restricted in his choice of working hours, due to a lack of information or a lack of mobility. If it is assumed that employers offer jobs with fixed numbers of hours, then the worker has to choose from a finite set of jobs, offering fixed numbers of hours. Since working zero hours is always possible, it will be treated as a special case. Let us assume that the market distribution of job offers is the same for all workers, such that the probability of a job offer, which involves $h_\ell(\neq 0)$ working hours is:

$$\Pr(\text{job offer } h=h_\ell) = p_\ell, \qquad \ell=1,\ldots,m \tag{6.7}$$

So we assume that there are m different values of working hours $h_\ell > 0$. And there is always the option of working zero hours. Then the labour supply decision becomes a discrete choice out of, let us say, N job offers, drawn from this market distribution of offers, and not working. If the number of job offers, N, approaches infinity, this model becomes a model without hours restrictions, see Appendix 6A. In that case the worker's behaviour can be described by a discrete choice model in which all possible values of hours are available:

$$h = h_j \quad \text{iff} \quad U(h_j,y_j) > U(h_k,y_k) \quad k=0,\ldots,m \ \text{ and } k\neq j \tag{6.8}$$

where U is specified by equations (6.1) and (6.6)
The index k covers the whole range of possible values of hours. However, if individuals face a limited choice of all job offers, then the index k only covers the range of job offers received and zero. One way to model this restricted choice problem is to write down all possible sets of job offers, with corresponding probabilities that an individual will get such a set of offers. Then the probability of observing h_j hours of work is the sum over all sets of the probability that h_j hours is preferred to all job offers in a specific set, times the probability of occurence of that set.

Although this way of modelling is appealing for its conceptual simplicity, it is computationally cumbersome. We will therefore use a different, although equivalent approach. In Appendix 6A the two methods are written out explicitly. The idea of the approach we use is that an individual is only observed to work h_j hours if he received at least one job offer h_j and if he preferred this job offer to all the other, different, offers he received and to not working. The individual is observed as a nonworker if he preferred zero hours to all job offers he received.

Given the values of the two random variables ε_h and ε_w it is possible to construct a set J_j consisting of all job offers preferred strictly less than $h=h_j$

$$J_j(\varepsilon_h,\varepsilon_w)=\{h_\ell:U(h_\ell,y(h_\ell);\varepsilon_h,\varepsilon_w)<U(h_j,y(h_j);\varepsilon_h,\varepsilon_w)\ ,\ \ell=1,\dots,m\} \tag{6.9}$$

Notice once more that h=0 is not treated as a job offer. The set $J_j \cup \{h_j\}$ contains all possible job offers an individual could have received if h_j is observed. If this individual would have received an offer that does not belong to $J_j \cup \{h_j\}$, he would have preferred that offer, and he would not have been observed to work h_j hours. Define Q_j as the probability that one draw out of the market distribution of job offers will yield an offer which is less preferred than the h_j chosen, i.e. the offer is in the set J_j

$$Q_j = \sum_{h_\ell \in J_j} p_\ell \tag{6.10}$$

Then the probability of observing $h=h_j$ $(h_j \neq 0)$ if N job offers have been received can be written as

$$\begin{aligned} R_j &= (Q_j + p_j)^N - Q_j^N \qquad &&\text{if } U(h_j,y(h_j);\varepsilon_h,\varepsilon_w)>U(0,y(0);\varepsilon_h,\varepsilon_w) \\ &= 0 &&\text{otherwise} \end{aligned} \tag{6.11}$$

The first line in equation (6.11) describes the probability that the individual only received offers which he preferred less than h_j and that at least one job offer was $h=h_j$. This is equivalent to saying that this individual drew N times a job offer out of $J_j \cup \{h_j\}$ (i.e. $(Q_j + p_j)^N$) but that he did not draw offers *only* out of J_j (Q_j^N). The second line in equation (6.11) says that if zero is preferred to h_j then the probability of observing h_j is zero, since zero is always available.

The probability of observing h=0 when N job offers are received is simply

$$R_0 = Q_0^N \tag{6.12}$$

Q_0^N is the probability that the N job offers are preferred less than h=0. In this model there are four sources of randomness, namely:

- ε_h, representing stochastic preferences
- ε_w, representing measurement errors in the wage equation
- N, the number of job offers
- the number of hours offered.

Recall that so far all formulas have been derived, conditional on the values of ε_h and ε_w. Removing the conditioning on ε_w is equivalent to taking into account the wage equation, and formulating a joint hours-wage probability. This is postponed to the next subsection.

The way to remove the conditioning on the unobserved taste parameter ε_h, is to integrate it out. In doing so, one should remember that the probability R_j is also conditional on the value of ε_h. Then the likelihood of observing $h=h_j$ hours given ε_w, can be written as

$$L(h=h_j|\varepsilon_w) = \int_{-\infty}^{\infty} \varphi(\varepsilon_h|\varepsilon_w)R_j(\varepsilon_h)\, d\varepsilon_h \tag{6.13}$$

where φ is the normal density function of ε_h given ε_w.

It is clear that J_j, the set of job offers less preferred than h_j, is a step function in ε_h, because only discrete values of hours are considered. Step changes occur at $\varepsilon_h=e_{jk}$, i.e. when ε_h takes on such a value that working h_j hours yields the same utility as working h_k hours. See the Appendix for the exact formula of e_{jk}. For values of ε_h between e_{jk} and e_{jk-1} the set $J_j(k)$ remains the same. $J_j(k)$ is defined as J_j (equation 6-.9) for $e_{jk-1}<\varepsilon_h< e_{jk}$. Switching from integrals to sums, we can rewrite (6.13) as follows

$$L(h=h_j|\varepsilon_w) = Pr(\varepsilon_h < e_{j0}|\varepsilon_w)\ R_j(0) + \sum_{k=1}^{m} Pr(e_{jk-1} < \varepsilon_h < e_{jk}|\varepsilon_w)\ R_j(k) + \tag{6.14}$$

$$Pr(\varepsilon_h > e_{jm}|\varepsilon_w)\ R_j(\text{rest})$$

where

$$R_j(0) = Pr(\ h=h_j|\ \varepsilon_h < e_{j0}, \varepsilon_w)$$

$$R_j(k) = Pr(\ h=h_j|\ e_{jk-1} < \varepsilon_h < e_{jk}, \varepsilon_w),\ k=1,\dots,m$$

$$R_j(\text{rest}) = Pr(\ h=h_j|\ \varepsilon_h > e_{jm}, \varepsilon_w)$$

For details on (6.14), the reader is referred to Appendix 6A.

Until now, we have taken the number of job offers, N, to be fixed. But as mentioned in the introduction, in order to be able to capture the possibility of involuntary unemployment, we make N stochastic. The only difference with the formulas above is that we have to take expectations with respect to the number of job offers

$$L'(h=h_j|\varepsilon_w) = \sum_{N=0}^{Nmax} L(h=h_j|\varepsilon_w,N)\cdot p(N) \tag{6.15}$$

where

L' is the likelihood when N is a random variable

L is given in equation (6.14)

Nmax is the maximum number of job offers

p is a discrete probability distribution.

Because $L(h=0|\varepsilon_w,N=0)=1$ (see formula (6.12)), equation (6.15) turns into

$$L'(h=0|\varepsilon_w) = p(0) + \sum_{N=1}^{Nmax} L(h=0|\varepsilon_w,N)\cdot p(N) \tag{6.16}$$

for nonworkers. So not working is either explained by the fact that an individual didn't receive any job offers at all (p(0)) or because he preferred not working while he received N>0 job offers (second term).

Because $L(h=h_j|\varepsilon_w,N=0)=0$ (see formula (6.11)), equation (6.15) can be rewritten as

$$L'(h=h_j|\varepsilon_w) = \sum_{N=1}^{Nmax} L(h=h_j|\varepsilon_w,N)\cdot p(N) \tag{6.17}$$

for workers.

With respect to identification of the model the following comments are in order. In the case of no hours restrictions, the parameters are known to be identifiable if the Z-vector in equation (6.3) contains one variable not in the X-vector in equation (6.1) (Moffitt (1984)). Adding hours restrictions makes it much more complex. We observe the joint sample of hourly wages and actual hours of work. This joint observed distribution is assumed to be generated by the distribution of budget constraints, the distribution of preferences and the distribution of job offers. Clearly, if we do not assume functional forms for the budget constraint, the utility function and the job offer distribution, we cannot identify the parameters in the model. The main identifying assumption is that the assumed functional form of the job offer distribution is different from the functional form of the preferred hours distribution. Moreover the job offer distribution does not depend on individual characteristics, while the preferred hours distribution does. Although we realize that the identifying assumptions are strong, we believe that the model as it is provides one sensible, among many possible, interpretation of labour market behaviour. The economics behind the hours restrictions model can be seen as follows. For many individuals in The Netherlands preferred hours seem to be lower than actual hours worked (Kapteyn and Woittiez (1989)). Part of the explanation of this fact can be found in mandatory overtime. But surely part of it is due to the fact that full time jobs are largely overrepresented and part time jobs are relatively scarce. This may be caused by fixed costs that may make employers favourable to full time jobs. And it might be impossible for the employer to lower the offered wage for part time jobs, since wages are tied to a certain type of job. Other interpretations of the job offer distributions could be unobserved wage variation, measurement error, fixed costs at the supply side of the labour market or unobserved preference variation. If unobserved preference variation is the explanation for the peaked hours distribution, then the random error term must be multinomial distributed. But it seems realistic to assume that unobserved preference variation is smooth and not peaked. And this would make the unobserved preference variation implausible. Fixed costs for the individual could explain the small portion of part-time jobs, since it generates a horizontal budget constraint at low hours of work (Moffitt (1984)). But it provides no explanation for the other spikes in the hours distribution at 20 and 32 hours per week. It could be just measurement error that makes actual hours differ from preferred hours. If the estimated job offer distribution would have been smooth, this might be a plausible interpretation. But we don't think that a peaked distribution could be sensibly explained as measurement error. Unobserved fringe benefits that are greater for full time jobs than for part time jobs could lead to few people working part time. But once again it cannot provide an explanation for the spikes at 20 and 32 hours per week. Finally one remark with respect to the identification of the parameter Nmax, the maximum number of job offers an individual receives. If Nmax=1 the model describes a take-it-or-leave-it decision. For large values of Nmax the job offer distribution (p's) and the number of job offers (N) introduce the same kind of flexibilty and identification becomes troublesome.

- Formulation of the joint wage-hours model

As yet the model has been derived conditional on ε_w. The removal of this conditioning amounts to adding the wage equation to the model and formulating the joint probability of observing h_j hours of work and the corresponding wage rate w. For workers the joint probability can be defined as

$$L'(h=h_j,w) = L'(h=h_j|\varepsilon_w)\ L''(\varepsilon_w=w - Z\psi - bh_j - ch_j^2) \tag{6.18}$$

The first term of this probability is given in equation (6.15). For nonworkers equation (6.18) has to be adapted, since for nonworkers the wage rate is not observed. Therefore, the unobservable wage rate must be integrated out, which results in a similar likelihood of observing h=0 as in equations (6.14) and (6.16)

$$L'(h=0) = p(0) + \sum_{N=1}^{Nmax} [\ Pr(\ u< u_{01})\ R_0(1) +$$

$$\sum_{k=2}^{m} \{\Pr(\ u_{0k-1} < u < u_{0k})\ R_0(k)\} + \tag{6.19}$$

$$\Pr(\ u > u_{0m})\ R_0(\text{rest})] \cdot p(N)$$

where $u_{0k} = e_{0k} + \gamma\varepsilon_w$ is the new limit of integration similar to e_{0k}. See Appendix 6A for further details.

6.3 Estimation Results

The model is described by equations (6.1), (6.3), (6.5) and (6.8); the exact specification of the likelihood function is found in Appendix 6A. The number of hours in this model is assumed to be a discrete variable with 4-hour intervals, i.e. 0, 4, 8 etc. Naturally the actual number of hours worked and not the preferred number of hours is used as the endogenous hours variable. For the hours variable is interpreted as the outcome of the interplay between preferred hours and demand side restrictions. In estimating the model net wages are used.

In Tables 6.1-6.4 estimation results are presented for males and females in the OSA-sample and in the SEP-sample, respectively. The first and third columns show results for the model without hours restrictions and with fixed wages (i.e. in the wage equation the wage is not dependent on hours of work). In principle, this is a standard hours-wage model with a linear budget constraint. The second and fourth columns of Tables 6.1-6.4 correspond with the model sketched above. We have classified hours into eight groups, with different probabilities of being drawn. These probabilities are indicated in the tables by p's. For example, the probability of receiving a job offer of 4 hours a week equals p_4. The number of job offers N is a random variable and is assumed to follow a binomial distribution $B(p_N, Nmax)$, where Nmax has been fixed at 15. In Tummers and Woittiez (1989) the model has been estimated for different values of Nmax. Their results show that the expected number of job offers ($p_N Nmax$) remains the same, and that the values of the log likelihoods are close for Nmax=10 and 15. The log likelihood value for Nmax=3 is lower. Apparently Nmax must be large enough to provide enough flexibility.

Let us now turn to the estimation results. Comparing the likelihoods of column one and two with each other, and of columns three and four, we can conclude that the joint hypothesis of the wage independence of hours of work and no hours restrictions can be rejected. In the model with hours restrictions the hours coefficients b and c are in general insignificant, but in Tummers and Woittiez it is shown that their joint effect is significant, although less so than in the model without hours restrictions.

One should bear in mind that in the case of the nonlinear budget constraint the labour supply equation is a cubic equation in h. Therefore we have to be very careful when we compare the values of the estimated parameters in this case with the values in the linear case. We can see that for all versions of the model, the female labour supply curve is forward bending (γ>0) and the male labour supply curve is inelastic. This can also be seen from Tables 6.5 and 6.6 which present wage elasticities. It is most interesting to note that all wage elasticities are lower in the extended model than in the standard model. See also Figures 6.1 to 6.4. Nonlabour income has a negative effect on hours of work for both males and females. Family size has a positive effect on the male labour supply and a negative effect on the female labour supply. Turning to the wage equation, we notice that education has a positive effect on wages and that wages increase with age until about 50 to 60 years.

The estimated value of p_N, the parameter in the job offer distribution, is higher for males than for females. Males in families in the OSA-sample and single males in the SEP-sample all receive the assumed maximum of 15 job offers, while single males in the OSA-sample and married males in the SEP-sample receive around 3 job offers. The reason for the big difference in the number of job offers received between the various groups can only be explained by an identification problem. This is probably due to the lack of variation in hours worked. Females receive on average about 3 or 4 job offers. Like for males, most job offers involve 40 hours per week or more. However, preferences are such that these offers will be rarely accepted. In Figures 6.5 through 6.8 hours distributions are drawn for each version of the model. From Figures 6.5 and 6.6 we observe that the models without hours restrictions do not predict the actual hours distribution very well (see also previous chapters). The models miss the peaks at 20, 32 and 40 hours. By including more flexibility through a nonlinear budget constraint and hours restrictions, both the underprediction of the nonworking and the overprediction of the inviduals working low hours are reduced.

It is interesting to take a closer look at how the simulated degree of nonparticipation has been achieved. Let us concentrate on married females in the OSA-sample. In Table 6.7 the actual, simulated, and preferred hours distribution generated by the model are presented as well as the job offer distribution. The third column suggests that only 7.7% of the females prefer not working to any other number of hours. The remaining 51.2% of the predicted nonparticipation must be due to restrictions on the demand side. A large number of women, 75.1%, prefer to work between 4 and 20 hours per week. But jobs requiring such low number of hours are rarely offered. This could for example be explained by the existence of fixed costs to the employer for each seperate employee.

Table 6.1 Estimation results for males, OSA[a)]

	single males		males in families	
	Standard	Extended	Standard	Extended Model
δ_{10}	38.7(1.7)	35.3(1.1)	41.8(0.9)	39.2(1.6)
β_1	-0.0001(0.0001)	-0.0002(0.0001)	-0.004(0.001)	-0.000001(ℓ.b.)
δ_{11}	4.9(3.1)	10.5(2.0)	0.9(0.7)	1.8(1.2)
γ_1	0.01(0.11)	0.01(ℓ.b.)	0.01(ℓ.b.)	0.01(ℓ.b.)
σ_h	9.1(0.7)	7.3(0.7)	7.2(0.2)	9.3(0.5)
a_1	-4.8(7.3)	-11.5(4.1)	-1.8(2.7)	3.4(4.6)
a_{22}	2.5(1.9)	2.8(1.2)	0.7(0.6)	0.7(0.6)
a_{23}	5.7(1.7)	5.5(1.1)	2.2(0.5)	2.1(0.5)
a_{24}	7.6(1.7)	6.6(1.0)	5.2(0.5)	5.1(0.5)
a_{31}	0.6(0.4)	0.6(0.2)	0.7(0.1)	0.8(0.1)
a_{32}	-0.005(0.005)	-0.005(0.003)	-0.007(0.002)	-0.009(0.002)
b	0(fixed)	0.3(0.1)	0(fixed)	-0.2(0.2)
c	0(fixed)	-0.003(0.002)	0(fixed)	0.0002(0.002)
σ_w	4.7(0.3)	4.3(0.3)	4.8(0.1)	4.6(0.1)
ρ_{hw}	-0.5(0.1)	-0.99(u.b.)	-0.3(0.04)	0.04(0.13)
p_{4-16}		0.03(0.02)		0.002(0.001)
p_{20}		0.01(0.01)		0.003(0.002)
$p_{24,28}$		0.02(0.01)		0.002(0.001)
p_{32}		0.04(0.02)		0.004(0.001)
p_{36}		0.02(0.01)		0.006(0.001)
p_{40}		0.33(0.05)		0.84(0.04)
p_{44}		0.20(0.04)		0.06(0.02)
pn		0.25(0.03)		0.99(u.b.)
log lik	-603.0	-544.0	-3977.8	-3646.4

a) Standard errors in parentheses
u.b. = upper bound, ℓ.b. = lower bound

Table 6.2 Estimation results for females, OSA[a)]

	single females		females in families	
	Standard	Extended	Standard	Extended Model
δ_{10}	25(fixed)	25(fixed)	8.6(9.7)	21.3(3.7)
β_1	-0.0010(0.0001)	-0.0008(0.0002)	-0.012(0.004)	-0.007(0.002)
δ_{11}	-2.9(3.2)	-2.5(3.9)	-48.6(3.2)	-23.9(4.6)
γ_1	1.1(0.1)	1.0(0.4)	4.6(0.7)	2.3(0.6)
σ_h	13.2(0.9)	7.0(3.3)	28.3(2.5)	10.0(2.5)
a_1	-2.6(4.5)	-0.6(5.4)	-4.2(2.8)	2.6(3.3)
a_{22}	0.3(1.4)	1.5(1.5)	0.2(0.5)	0.2(0.5)
a_{23}	0.7(1.1)	1.5(1.2)	1.7(0.4)	2.0(0.5)
a_{24}	2.2(1.2)	3.0(1.3)	4.4(0.6)	5.4(0.7)
a_{31}	0.8(0.2)	0.6(0.3)	0.9(0.1)	0.7(0.2)
a_{32}	-0.009(0.003)	-0.007(0.003)	-0.012(0.002)	-0.010(0.002)
b	0(fixed)	0.2(0.2)	0(fixed)	-0.1(0.1)
c	0(fixed)	-0.005(0.002)	0(fixed)	0.000(0.001)
σ_w	4.3(0.3)	4.0(0.3)	3.9(0.2)	3.9(0.2)
ρ_{hw}	-0.5(0.1)	-0.4(0.2)	-0.6(0.1)	-0.6(0.1)
p_{4-16}		0.01(0.005)		0.01(0.002)
p_{20}		0.06(0.03)		0.02(0.01)
$p_{24,28}$		0.01(0.005)		0.01(0.003)
p_{32}		0.02(0.01)		0.04(0.01)
p_{36}		0.03(0.01)		0.03(0.01)
p_{40}		0.23(0.04)		0.25(0.06)
p_{44}		0.18(0.05)		0.12(0.04)
pn		0.26(0.04)		0.31(0.07)
log lik	-710.0	-633.1	-2109.9	-1961.2

a) Standard errors in parentheses

Table 6.3 Estimation results for males, SEP[a)]

	single males		males in families	
	Standard	Extended	Standard	Extended Model
δ_{10}	36.1(1.6)	31.3(1.7)	39.0(1.2)	34.1(1.3)
β_1	-0.012(0.006)	-0.000001(u.b)	-0.011(0.001)	-0.005(0.001)
δ_{11}	4.7(3.7)	3.5(2.7)	2.3(0.9)	2.9(0.9)
γ_1	0.01(ℓ.b.)	0.01(ℓ.b.)	0.08(0.03)	0.01(ℓ.b.)
σ_h	13.5(0.9)	10.6(1.0)	8.7(0.2)	3.2(0.5)
a_1	3.0(4.3)	6.4(5.3)	1.8(2.5)	1.1(4.2)
a_{22}	1.9(1.0)	2.0(1.0)	1.1(0.5)	1.0(0.5)
a_{23}	3.1(0.9)	3.2(0.9)	2.3(0.4)	2.2(0.4)
a_{24}	4.7(1.0)	4.7(1.0)	6.5(0.5)	6.0(0.5)
a_{31}	0.4(0.2)	0.4(0.2)	0.4(0.1)	0.4(0.1)
a_{32}	-0.004(0.003)	-0.004(0.003)	-0.004(0.001)	-0.003(0.001)
b	0(fixed)	0.04(0.19)	0(fixed)	0.1(0.1)
c	0(fixed)	-0.004(0.003)	0(fixed)	-0.002(0.002)
σ_w	3.4(0.3)	3.4(0.3)	4.1(0.1)	4.1(0.1)
ρ_{hw}	-0.4(0.1)	0.4(0.2)	-0.2(0.1)	-0.99(ℓ.b.)
p_{4-16}		0.001(0.006)		0.002(0.002)
p_{20}		0.012(0.098)		0.010(0.007)
$p_{24,28}$		0.002(0.016)		0.009(0.003)
p_{32}		0.004(0.037)		0.012(0.003)
p_{36}		0.005(0.044)		0.006(0.002)
p_{40}		0.76(2.03)		0.384(0.017)
p_{44}		0.07(0.59)		0.114(0.012)
pn		0.99(8.4)		0.20(0.01)
log lik	-622.3	-502.4	-4496.8	-3732.9

a) Standard errors in parentheses

u.b. = upper bound, ℓ.b. = lower bound

Table 6.4 Estimation results for females, SEP[a)]

	single females		females in families	
	Standard	Extended	Standard	Extended Model
δ_{10}	-64.4(44.2)	-28.8(22.1)	-27.5(18.6)	5.0(3.8)
β_1	-0.05(0.01)	-0.01(0.01)	-0.005(0.004)	-0.00001(0.002)
δ_{11}	-4.1(4.6)	-3.4(2.8)	-46.4(3.4)	-15.7(2.5)
γ_1	7.9(3.4)	4.3(1.7)	5.9(1.3)	2.2(0.1)
σ_h	29.7(10.8)	15.3(5.8)	40.4(5.9)	12.3(1.0)
a_1	1.8(4.6)	2.4(6.0)	-3.2(3.8)	4.0(2.6)
a_{22}	0.4(0.5)	-0.1(0.4)	-0.3(0.4)	-0.1(0.4)
a_{23}	0.9(0.6)	0.3(0.4)	1.1(0.4)	1.5(0.4)
a_{24}	1.1(0.7)	0.6(0.5)	2.9(0.6)	3.1(0.6)
a_{31}	0.6(0.2)	0.4(0.3)	0.9(0.2)	0.6(0.1)
a_{32}	-0.007(0.003)	-0.005(0.004)	-0.012(0.003)	-0.008(0.002)
b	0(fixed)	0.11(0.05)	0(fixed)	-0.18(0.05)
c	0(fixed)	-0.0008(0.0007)	0(fixed)	0.003(0.001)
σ_w	3.0(0.3)	3.0(0.3)	4.8(0.2)	4.6(0.2)
ρ_{hw}	-0.9(0.1)	-0.98(0.03)	-0.8(0.1)	-0.9(0.02)
p_{4-16}		0.01(0.01)		0.02(0.01)
p_{20}		0.03(0.02)		0.03(0.01)
$p_{24,28}$		0.01(0.01)		0.02(0.01)
p_{32}		0.04(0.02)		0.04(0.01)
p_{36}		0.02(0.01)		0.04(0.01)
p_{40}		0.34(0.11)		0.53(0.09)
p_{44}		0.08(0.05)		0.05(0.03)
pn		0.37(0.10)		0.17(0.04)
log lik	-447.6	-376.5	-2138.4	-2021.9

a) Standard errors in parentheses

Table 6.5 Wage elasticities, standard model[a)]

	OSA		SEP	
	singles	families	singles	families
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.01	0.00	0.00	0.03
	(0.19)	(0.00)	(0.00)	(0.02)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.43	3.35	3.48	7.09
	(0.04)	(1.74)	(3.15)	(7.08)

a) Standard errors in parentheses

Table 6.6 Wage elasticities, extended model[a)]

	OSA		SEP	
	singles	families	singles	families
$\frac{\partial h_m}{\partial w_m}\frac{w_m}{h_m}$	0.01	0.00	0.00	0.00
	(0.01)	(0.00)	(0.00)	(0.00)
$\frac{\partial h_f}{\partial w_f}\frac{w_f}{h_f}$	0.27	1.85	2.76	1.84
	(0.17)	(0.85)	(2.32)	(0.34)

a) Standard errors in parentheses

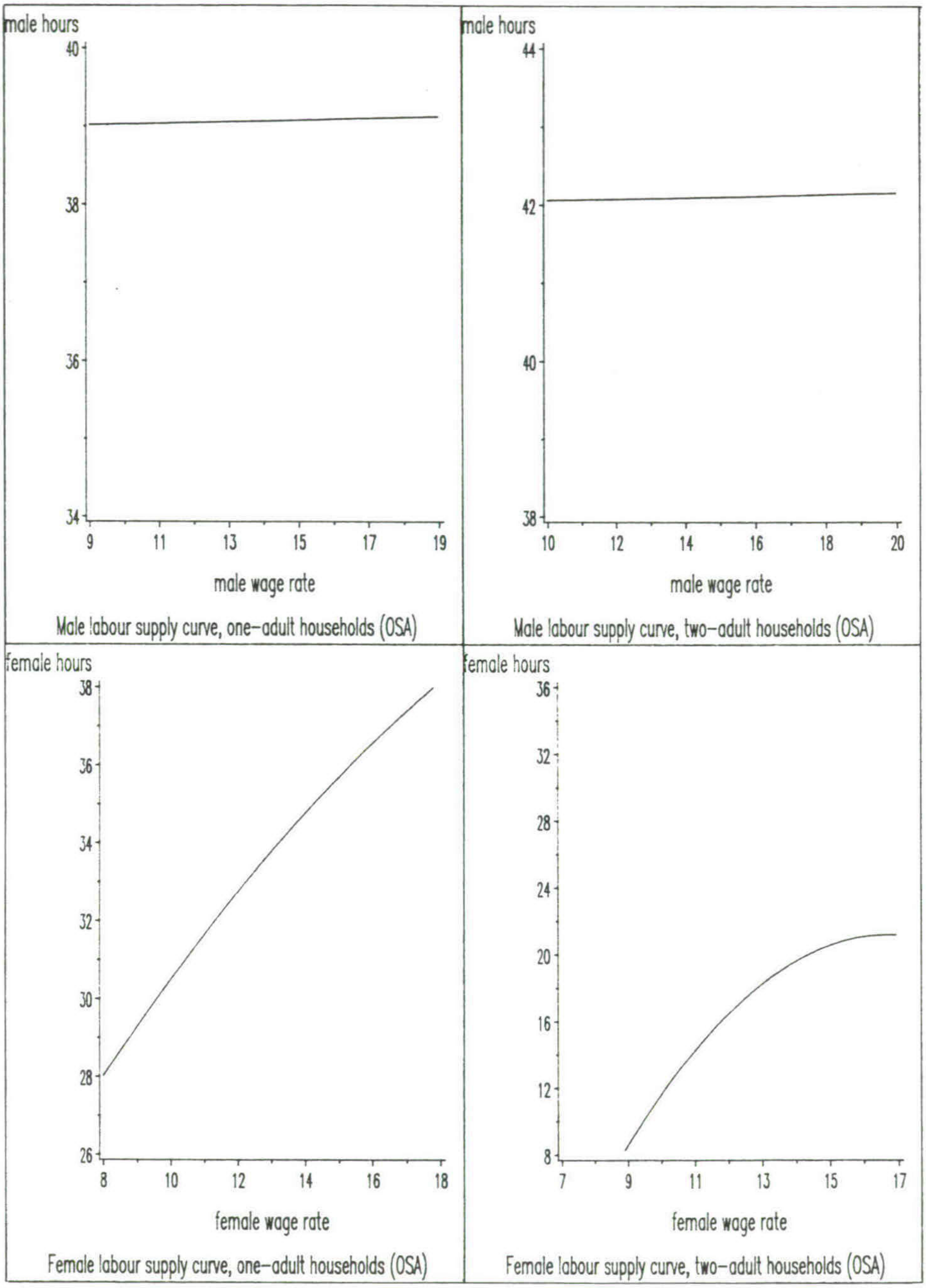


Figure 6.1 Labour supply curves, standard model (OSA)

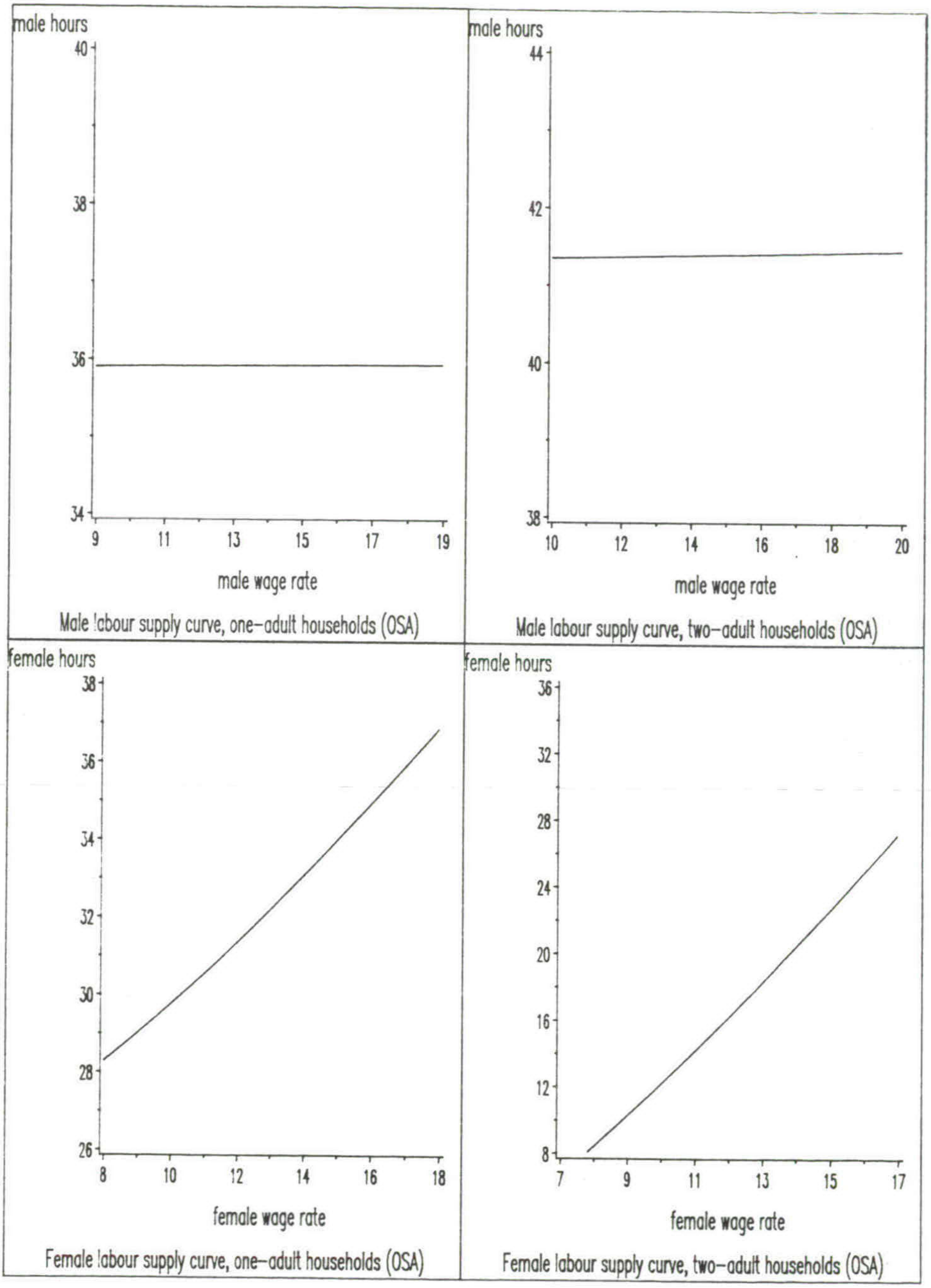


Figure 6.2 Labour supply curves, extended model (OSA)

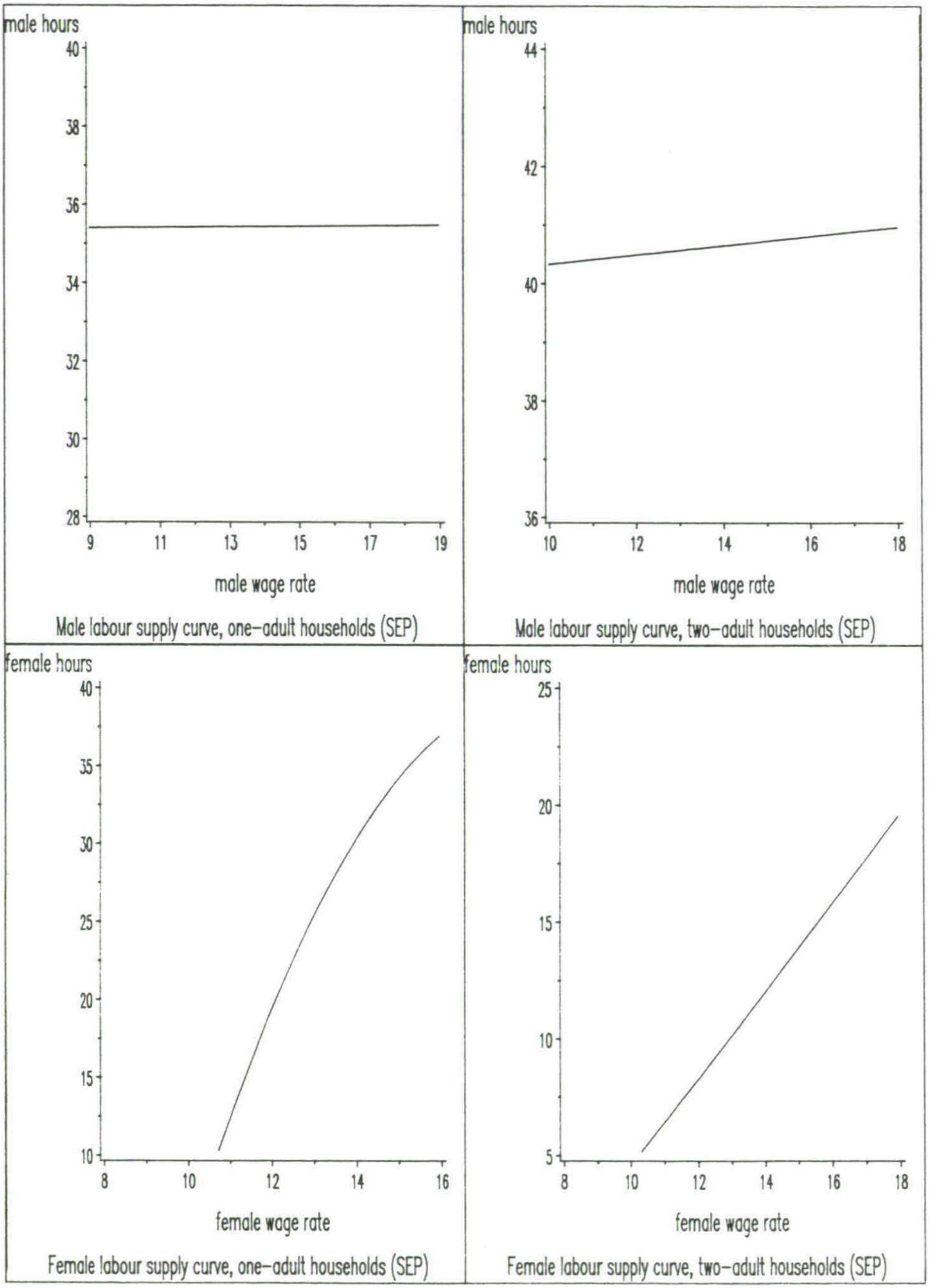


Figure 6.3 Labour supply curves, standard model (SEP)

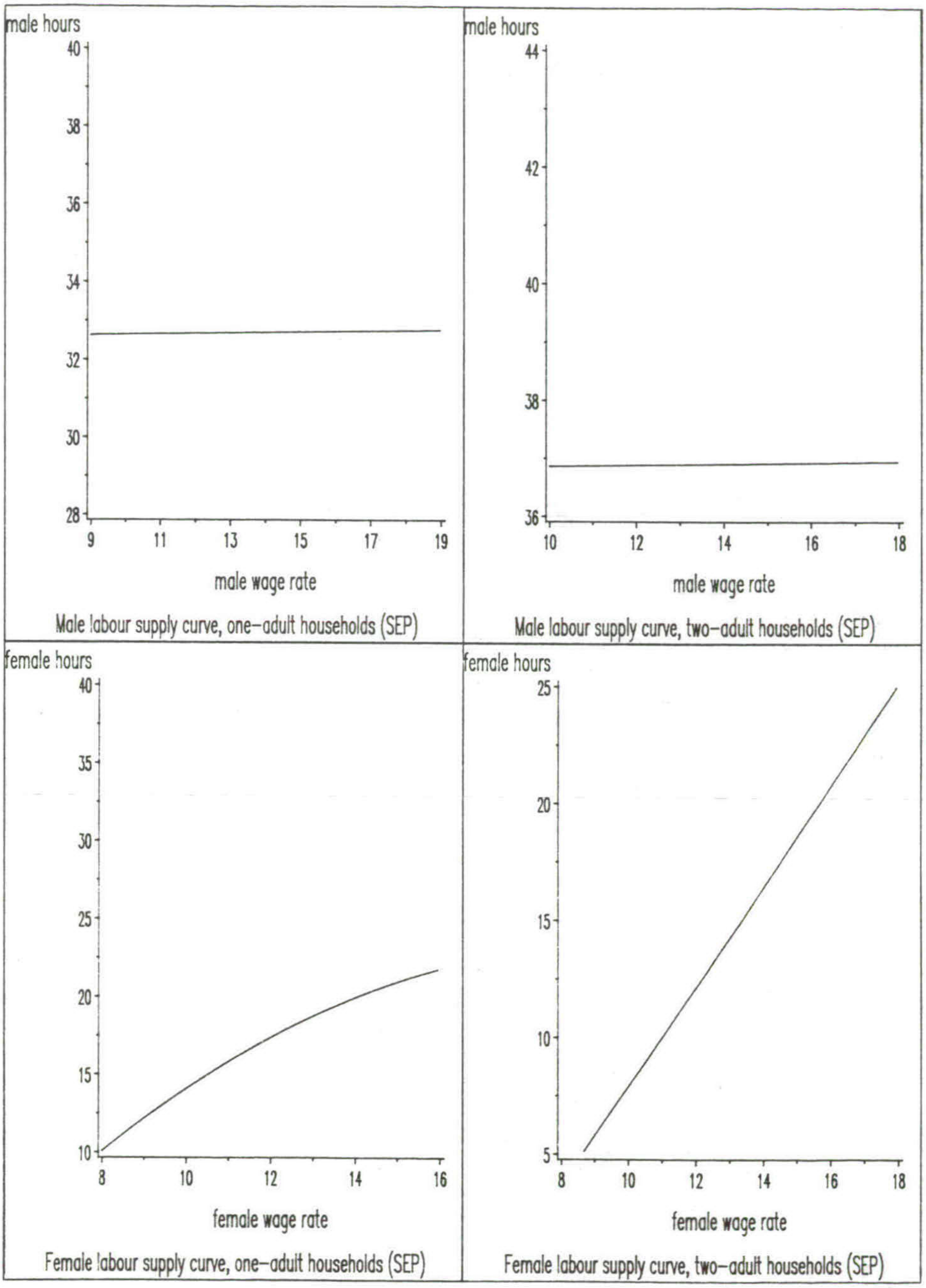


Figure 6.4 Labour supply curves, extended model (SEP)

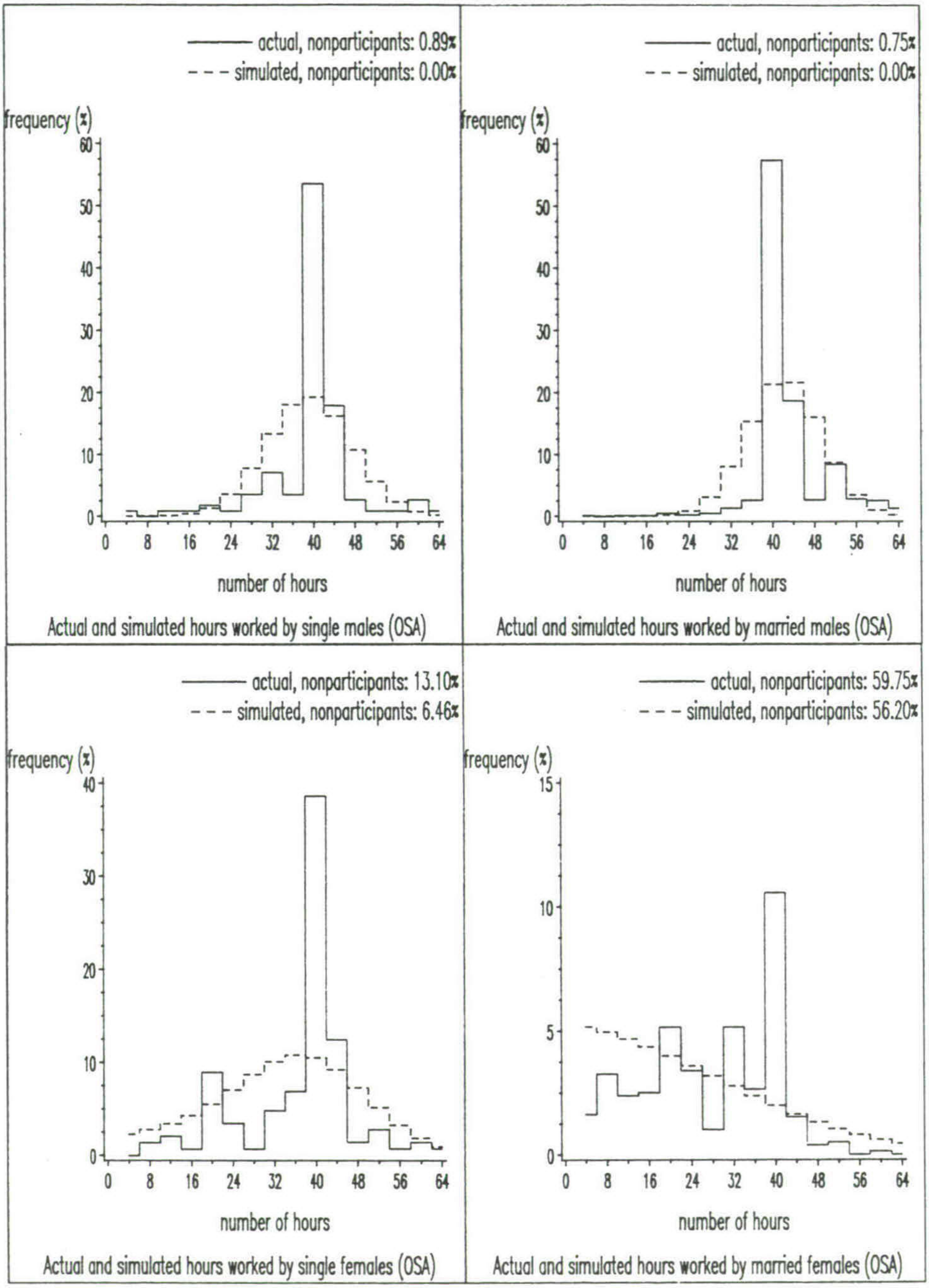


Figure 6.5 Hours distributions, standard model (OSA)

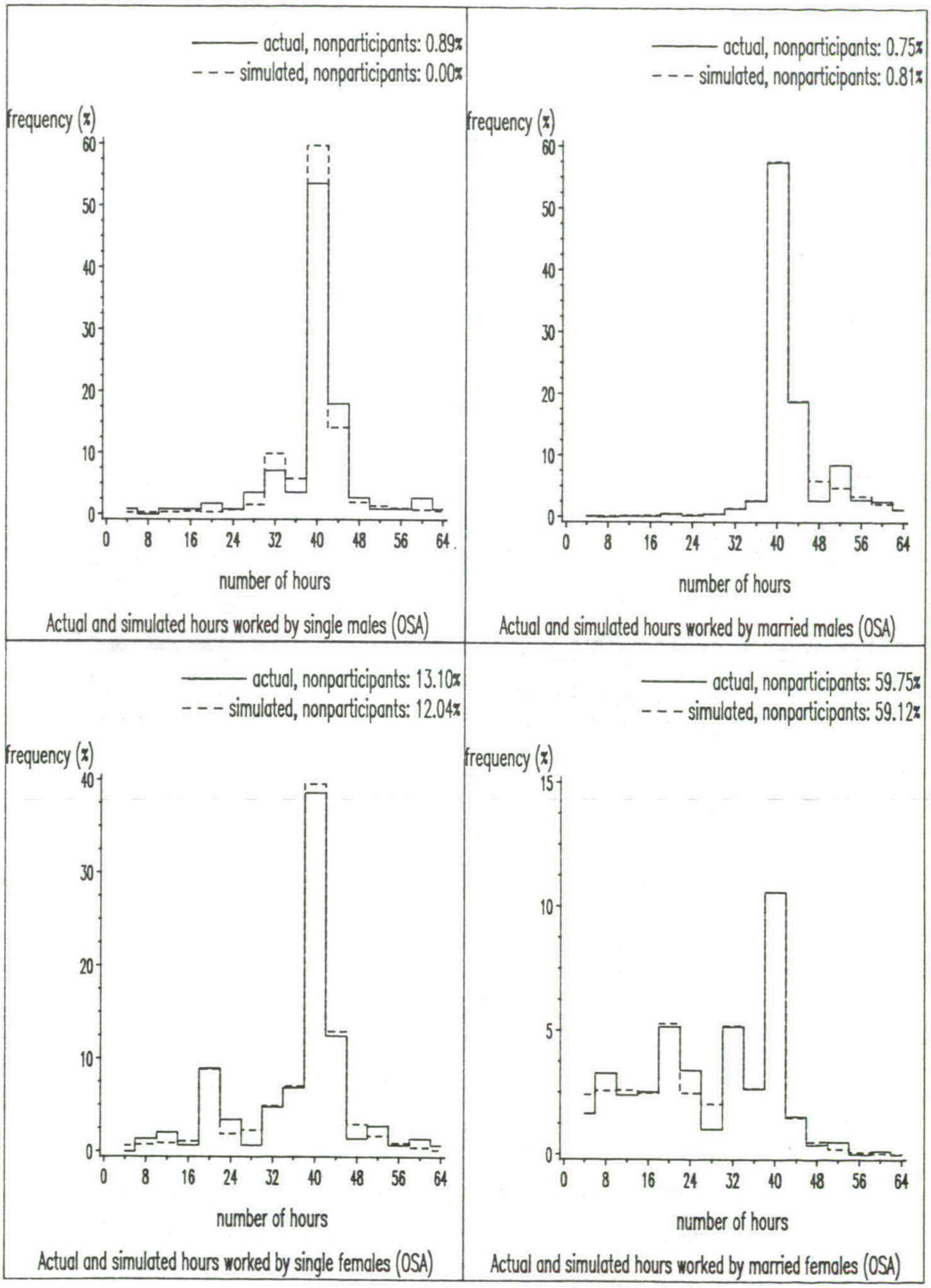


Figure 6.6 Hours distributions, extended model (OSA)

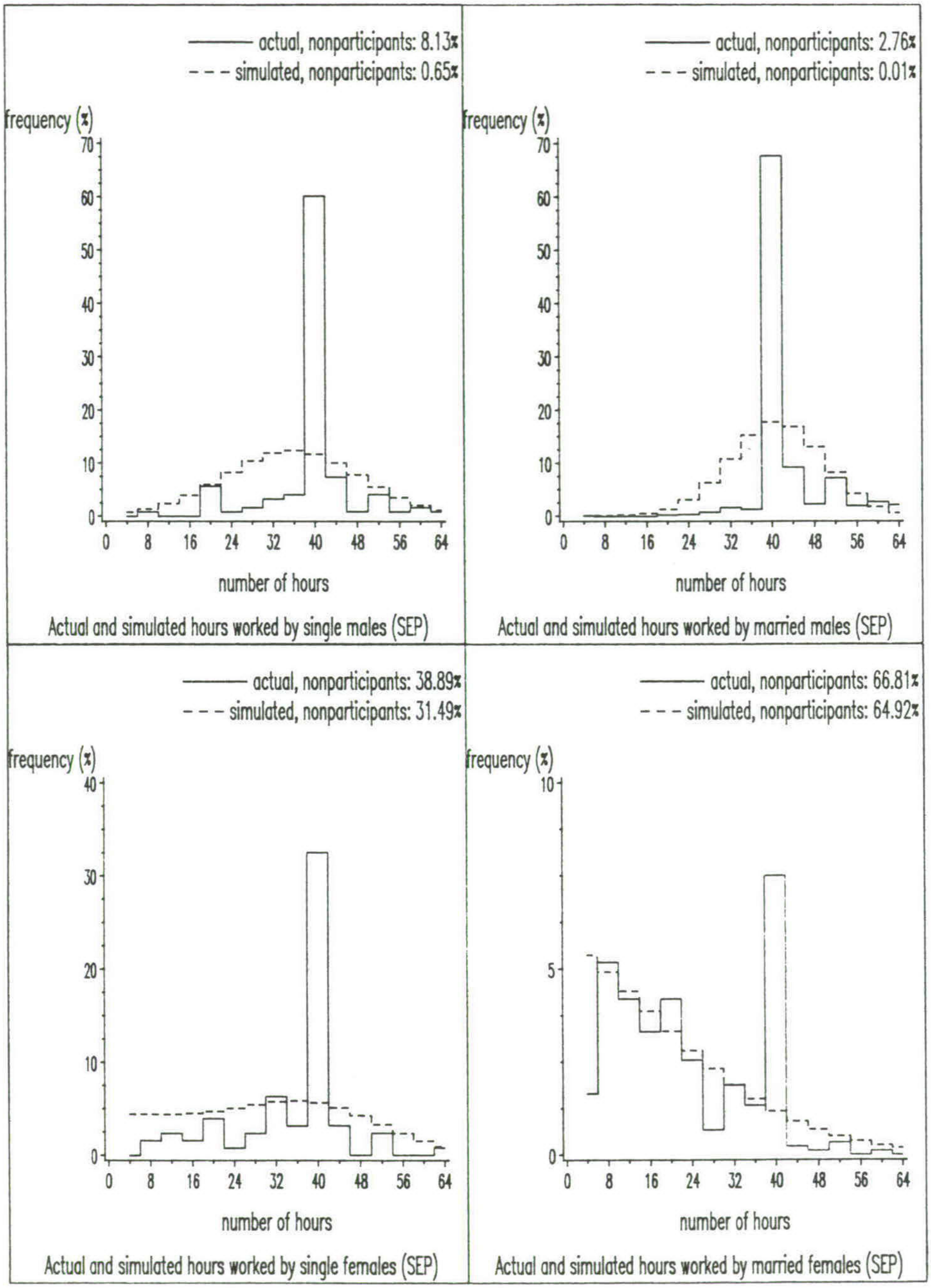


Figure 6.7 <u>Hours distributions, standard model (SEP)</u>

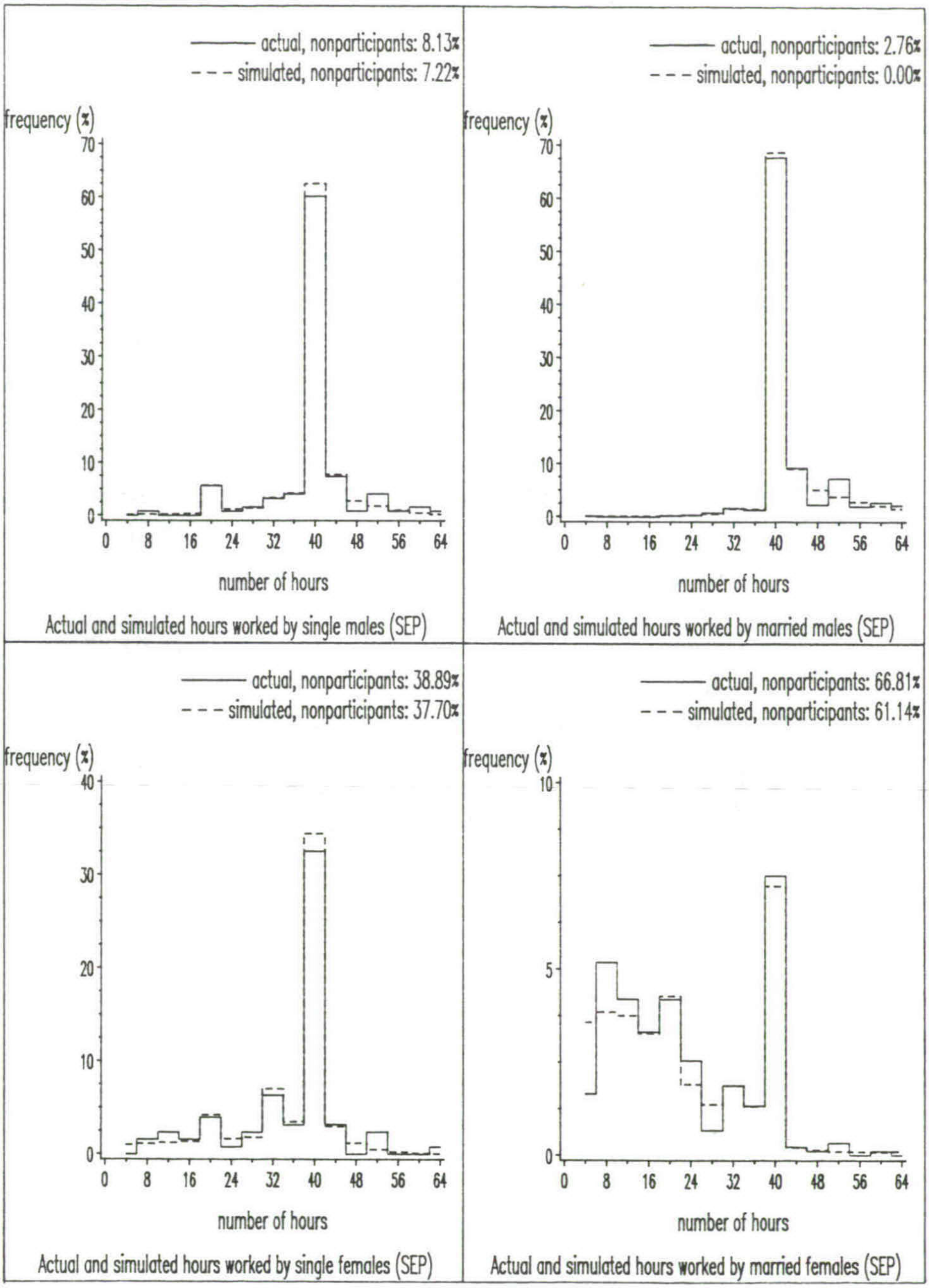


Figure 6.8 Hours distributions, extended model (SEP)

Table 6.7 Actual, simulated and preferred hours distributions, probabilities x 100, married females, OSA

hours	actual	simulated	preferred	offered
0	59.75	59.12	7.73	
4	1.64	2.40	10.19	0.85
8	3.27	2.56	16.00	0.85
12	2.39	2.58	18.51	0.85
16	2.52	2.47	16.85	0.85
20	5.16	5.29	13.48	2.20
24	3.40	2.47	9.45	1.10
28	1.01	2.03	5.15	1.10
32	5.16	5.19	2.00	3.76
36	2.64	2.66	0.54	2.76
40	10.57	10.57	0.09	25.36
44	1.51	1.47	0.01	11.92
48	0.38	0.50	0.00	9.72
52	0.50	0.21	0.00	9.72
56	0.00	0.08	0.00	9.72
60	0.13	0.03	0.00	9.72
64	0.00	0.00	0.00	9.72

6.4 Conclusion

A full simultaneous model of labour supply and wage determination with hours restrictions is estimated. That is, the possibility that an individual is faced with the limited availability of jobs with different, distinct, numbers of hours has been incorporated. Moreover, it has been taken into account that the wage rate may be dependent on the number of working hours. This leads to a nonlinear budget constraint.

Incorporating hours restrictions and hours dependent wages into the standard labour supply-wage model produces a better approximation of the actual hours distribution. This model may be a natural first step towards confronting labour supply with the demand side.

One of the limitations of this model is that it is static. In the future the model could be extended with job offers that arrive consecutively. Another drawback of this study is that the hours restrictions are imposed by a distribution of job offers, common to all individuals. Further research requires a more structural specification in this respect, to account for differences in employment opportunities across individuals. Furthermore, the social security and welfare system could be considered explicitly in describing the budget constraint.

Appendix 6A The expression for e_{jk} and the likelihood function

Let us first give the exact specification of the e_{jk}'s, i.e. the value of ε_h for which utility in h_j equals utility in h_k. After that we will give an alternative and probably less efficient formulation of our model. Then the entire model we have estimated will be given, and we will show that the standard model is a special case of the extended model.

The values e_{jk} follow from equating utility between points h_j and h_k, satisfying the budget constraint.

$$U(h_j,y_j) = U(h_k,y_k) \tag{6A.1}$$

Using equation (6.1) and substituting equation (6.6) gives

$$\log(\gamma-\beta h_j)+\frac{\beta(h_j-X\delta-e_{jk}-\beta y_j)}{(\gamma-\beta h_j)}=\log(\gamma-\beta h_k)+\frac{\beta(h_k-X\delta-e_{jk}-\beta y_k)}{(\gamma-\beta h_k)} \tag{6A.2}$$

where $\quad y_j = h_j Z\psi + bh_j^2 + ch_j^3 + I + h_j\varepsilon_w$

$$y_k = h_k Z\psi + bh_k^2 + ch_k^3 + I + h_k\varepsilon_w$$

Simple calculations give the solution for e_{jk}

$$e_{jk} = \frac{(\gamma-\beta h_j)(\gamma-\beta h_k)}{\beta^2(h_k-h_j)}\log((\gamma - \beta h_k)/(\gamma - \beta h_j)) +$$

$$\frac{(\beta h_j-\gamma)(bh_k^2+ch_k^3) - (\beta h_k-\gamma)(bh_j^2+ch_j^3)}{(h_k-h_j)} + \tag{6A.3}$$

$$\gamma/\beta - X\delta - \beta I - \gamma Z\psi - \gamma\varepsilon_w$$

$$= u_{jk} - \gamma\varepsilon_w \tag{6A.4}$$

In Moffitt (1984) a general rule is derived for which $h=h_j$ is preferred to all other, discrete, numbers of hours

$$\max_{k<j} e_{jk} < \varepsilon_h < \min_{k>j} e_{jk} \quad \forall\, k \neq j \tag{6A.5}$$

Rule (6A.5) expresses a choice $h=h_j$ in an appropriate range of values of the unobserved tastes for work, ε_h. A higher value of ε_h corresponds to a greater taste for work and a lower value of ε_h to a lesser taste for work (see equation (6.2)). Then the rule says that the value of ε_h has to be higher than all those values e_{jk} equating utility between the choice h_j and lower number of working hours ($h_{kl}<h_j$) and has to be lower than all those values e_{jk} equating utility between the choice h_j and higher number of working hours ($h_{kh}>h_j$). This means that the indifference curve for which the choice (h_j,y_j) results needs to be flatter (i.e. higher ε_h) in point (h_j,y_j) than those indifference curves connecting point (h_j,y_j) with points (h_{kl},y_{kl}) (lower number of working hours) and needs to be steeper (i.e. lower ε_h in point (h_j,y_j) than those indifference curves connecting point (h_j,y_j) with points (h_{kh},y_{kh}) (higher number of working hours). This is illustrated in Figure 6A.1.

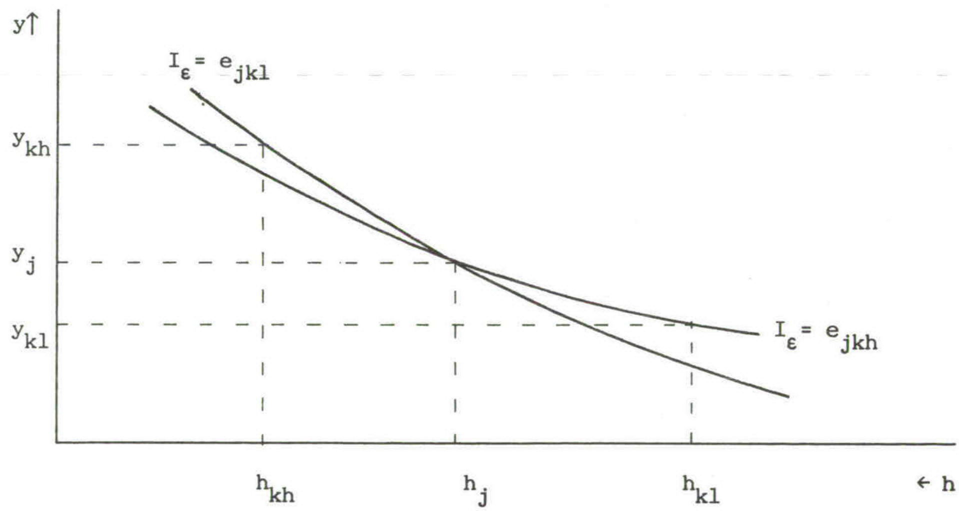


Figure 6A.1 The choice h_j if $\max_{k<j} e_{jk} < \varepsilon_h < \min_{k>j} e_{jk}$

Having this rule for ε_h and ε_h being normally distributed we can define probabilities for choice h_j if the choice set consists of all possible numbers of working hours h_k, $k=0,\dots,m$.

But if there is a limited number of job offers, so that individuals cannot choose freely their optimum number of hours, we have to write down all possible sets of job offers, and their probability of occurring. In writing down the probability of observing h_j hours, we are only interested in the sets with at least one offer h_j. There are $S = \sum_{i=1}^{N-1} \binom{m-1}{i} + 1$ number of these sets, to be called $V_j(s)$. Remember that N is the number of job offers and m is the number of possible positive distinct hours. Then the probability of observing h_j hours in the set $V_j(s)$ is the probability of occurence of the set $V_j(s)$ times the probability that h_j is preferred to all other numbers of hours in the set $V_j(s)$. The likelihood of observing $h=h_j$ hours, given ε_w, is the sum of the probability of observing h_j in the set $V_j(s)$ over all s

$$L(h=h_j|\varepsilon_w) = \sum_{s=1}^{S} \Pr[\max_{k<j} e_{jk} < \varepsilon_h < \min_{k>j} e_{jk} | h_k \epsilon V_j(s), \varepsilon_w] \Pr[V_j = V_j(s)] \qquad (6A.6)$$

We can see that in this way of modelling, determination of the probability of observing h_j is the same as looking for the appropriate range of values of ε_h, for all possible sets containing the offer $h=h_j$.

Instead of this formulation, we could also look for an appropriate set, for all possible values of ε_h. This is nothing else than changing the order of integration. We have now come to the formulation of the model we have estimated, except for the fact that the number of job offers is still fixed

$$\begin{aligned} L(h=h_j|\varepsilon_w) = {} & \Pr(h=h_j|\varepsilon_h < e_{j0}, \varepsilon_w) \Pr(\varepsilon_h < e_{j0}|\varepsilon_w) + \\ & \sum_{k=1}^{m} \Pr(h=h_j|e_{jk-1} < \varepsilon_h < e_{jk}, \varepsilon_w) \Pr(e_{jk-1} < \varepsilon_h < e_{jk}|\varepsilon_w) \qquad (6A.7) \\ & + \Pr(h=h_j|\ \varepsilon_h > e_{jm}, \varepsilon_w) \Pr(\varepsilon_h > e_{jm}|\varepsilon_w) \end{aligned}$$

Remember that

$$\Pr(h=h_j | e_{jk-1} < \varepsilon_h < e_{jk}, \varepsilon_w) = R_j(k)$$

where $R_j(k) = (Q_j + p_j)^N - Q_j^N \qquad$ if $U(h_j, y(h_j); \varepsilon_h, \varepsilon_w) > U(0, y(0); \varepsilon_h, \varepsilon_w)$

$= 0 \qquad$ otherwise

Implicit in this probability is the assumption that the individual is still a utility maximising person, because the set J_j (6.9) generating the probability Q_j only encloses job offers less preferred than the revealed choice h_j. Equation (6A.7) can be written explicitly as

For all h≠0 and h≠h_m

$$\begin{aligned} L(h=h_j | \varepsilon_w) = & \sum_{k=1}^{j-1} \{[\Phi(e_{jk}) - \Phi(e_{jk-1})]\, R_j(k)\} + \\ & [\Phi(e_{jj+1}) - \Phi(e_{jj-1})]\, R_j(j+1) + \\ & \sum_{k=j+2}^{m} \{[\Phi(e_{jk}) - \Phi(e_{jk-1})]\, R_j(k)\} + \\ & [1 - \Phi(e_{jm})]\, R_j(\text{rest}) \end{aligned} \tag{6A.8}$$

For h=0:

$$\begin{aligned} L(h=0 | \varepsilon_w) = & \Phi(e_{01})\, R_0(1) + \\ & \sum_{k=2}^{m} \{[\Phi(e_{0k}) - \Phi(e_{0k-1})]\, R_0(k)\} + \\ & [1 - \Phi(e_{0m})]\, R_0(\text{rest}) \end{aligned} \tag{6A.9}$$

For h=h_m:

$$L(h=h_m | \varepsilon_w) = \sum_{k=1}^{m-1} \{[\Phi(e_{mk}) - \Phi(e_{mk-1})]\, R_m(k)\} + \tag{6A.10}$$

$$[1 - \Phi(e_{mm-1})]\ R_m(m)$$

where Φ is the cumulative normal distribution function of ε_h conditional on ε_w and R_j(rest) is defined as in equation (6.13) where ε_h takes a value larger than e_{jm}.

A few remarks have to be made with respect to equations (6A.8)-(6A.10). First, the values e_{jk} have to be monotonically increasing in k. This will be the case if the budget constraint is linear, but if it is nonlinear this need not be the case. Then they have to be sorted, before the summing in equation (6A.8)-(6A.10) takes place. Second, the summation is split into two parts by the third term. The reason for this is that e_{jj} is not defined. The ranges $e_{jj-1}<\varepsilon_h< e_{jj}$ and $e_{jj}<\varepsilon_h< e_{jj+1}$ are combined in $e_{jj-1}<\varepsilon_h< e_{jj+1}$. The last term in equation (6A.8) ($\varepsilon_h>e_{jm}$) takes into account the right tail of the distribution of the unobserved ε_h. For workers the left tail ($\varepsilon_h<e_{j0}$) is not included in the summing because not working always belongs to the choice set and therefore working zero hours needs to be less preferred for a worker. So by rule (6A.5) the unobserved ε_h has to be greater than e_{j0}.

Furthermore, in the joint hours-wage model we have the difficulty that we do not observe the wage rate for nonworkers. In practice this means that the only term that we can use to define probabilities is $u_{0k} = e_{0k} + \gamma\varepsilon_w$ (see also equation (6A.3)). But given this joint unobserved effect, we are not able to evaluate utility. For in the utility function we find the expression $e_{0k} - \beta h_k\varepsilon_w$, and we only know $e_{0k} + \gamma\varepsilon_w$. We have solved this problem by using the fact that we need not know utility at h_k, but we only have to compare it with utility at h=0. We made use of the aforementioned rule

$$h=h_j \quad \text{iff} \quad \max_{k<j} e_{jk}<\varepsilon_h< \min_{k>j} e_{jk} \tag{6A.11}$$

Using $u_{jk} = e_{jk} + \gamma\varepsilon_w$ this is equivalent with

$$h=h_j \quad \text{iff} \quad \max_{k<j} u_{jk}<\varepsilon_h + \gamma\varepsilon_w< \min_{k>j} u_{jk} \tag{6A.12}$$

For nonworkers this turns into

$$h=0 \quad \text{iff} \quad u< \min_{k>0} u_{0k} \tag{6A.13}$$

From rule (6A.10) we know that if u is less than u_{0k}, then h_k is not preferred to 0 and therefore belongs to the set J_0

$$J_0 = \{ h_\ell : u<u_{0\ell},\ \ \ell=1,\dots,m \} \tag{6A.14}$$

So instead of comparing utilities to determine the set J, we compare the value $u = \varepsilon_h + \gamma\varepsilon_w$ with values $u_{0\ell}$.

To prove that the model without hours restrictions is a special case of the extended model if N is fixed, it is sufficient to show that the probability of observing $h=h_j$ is equal to one for a particular range of ε_h, if the number of job offers tends to infinity. The crucial expression is the probability of observing $h=h_j$ (see equation (6.10))

$$R_j = (Q_j + p_j)^N - Q_j^N \tag{6A.15}$$

Because Q_j is always smaller than one, R_j tends to one, for N tending to infinity, if $Q_j + p_j$ equals one. The probability $Q_j + p_j$ will only be equal to one for the workers if ε_h falls in a particular range. In all the other cases $Q_j + p_j$ will be less than one and R_j will go to zero. Similarly, there is a range for u such that Q_0 will equal one. We know those particular ranges for ε_h and u from equation (6A.5) and (6A.13). These values for ε_h and u, which are such that working h_j hours is more preferred to working h_k hours, $\forall$ k≠j show up in equations (6A.8) and (6.18)

$$[\Phi(e_{jj+1}) - \Phi(e_{jj-1})]\ R_j(j+1)$$

and

$$\Phi(u_{01})\ R_0(1)$$

So the model with fixed N converges to the model without hours constraints in Moffitt (1984). If N is a random variable, it must have a degenerate limiting distribution in infinity in order to attain an equivalent result.

7 Comparison of several models

7.1 Introduction

In Chapters 2 through 6 several models have been developed and estimated. The estimation results are based on two different datasets: the OSA- and the SEP-sample. In this chapter we will compare the estimation results in several ways. The comparison will not only deal with the various models but also with the two different datasets. With respect to the data it should be noted that both samples are supposed to be drawn from the same subset of the Dutch population in 1985, namely those individuals who are between 18 and 65 years old and who are able to work. We start with a test of this hypothesis (concentrating on the hours variables only). To that end we use a Pearson χ^2 test and two distribution free tests (Wilcoxon's test and the Kolmogorov-Smirnov test). The last two tests are nonparametric in the sense that they do not depend on the functional form of distribution functions. That is we test whether two distributions are equal, without specifying these distributions. In Section 7.2 the three two-sample tests are discussed and the test results are presented. We will also present some sample statistics such as means and standard deviations. Keeping the conclusions from Section 7.2 in mind we will in Section 7.3 compare the labour supply wage elasticities generated by one specific model that was estimated on the two data sets.

Besides looking at differences between elasticities corresponding with one specific model estimated on different data sets, we also compare different models estimated on the same data set. In each chapter we have presented a standard model and a so-called extended model, which is some generalization of the standard model. If the models are nested, we can test by means of a likelihood ratio test whether the restrictions imposed in the standard model hold.

We are also interested in how well the model simulations fit the actual data. To be able to say something about the fit of the various models we use χ^2 goodness-of-fit tests. When the parameters of the hypothesized distribution have been estimated by maximum likelihood, the conventional goodness-of-fit statistic no longer has a chi-squared distribution. In Heckman (1984) a new goodness-of-fit statistic is proposed which has a well defined chi-squared distribution. This statistic as well as the likelihood ratio statistic can be used to check for model misspecification. In Section 7.3 the various models estimated on two different data sets are evaluated by means of the aforementioned statistics. Finally in Section 7.4 conclusions will be drawn.

7.2 Two-sample tests

In this section we will perform various tests of the equality of the two populations from which the OSA- and SEP-samples were drawn. The problem we consider is the following

Let the scalars $X_1,\dots,X_m$ denote a random sample of size m from a cumulative density function (c.d.f.) $F_X(.)$ with corresponding density function $f_X(.)$ and let the scalars $Y_1,\dots,Y_n$ denote a random sample of size n from a cumulative density function $F_Y(.)$ with corresponding density function $f_Y(.)$. Further assume that the observations from $F_X(.)$ are independent of the observations from $F_Y(.)$. We want to test H_0: $F_X(z) = F_Y(z)$ for all z versus H_1: $F_X(z) \neq F_Y(z)$ for at least one value of z. Several tests are available. We first concentrate on two distribution free tests, namely Wilcoxon's test (also called the Mann-Whitney test or rank sum test) and the two-sample Kolmogorov Smirnov test.

- Wilcoxon's Test

Given two random samples $X_1,\dots, X_m$ and $Y_1,\dots, Y_n$ one arranges the m+n observations in ascending order and then replaces the smallest observation by 1, the next by 2 and the largest by m+n. These integers are called the ranks of the observations. The test is based on T_X, the sum of the ranks of the m X-values. (Since $T_X+T_Y= \sum_{j=1}^{m+n} j$ is constant the test could as well be based on T_Y.) Mann and Whitney (1947) have shown that under H_0 T_X is approximately normally distributed for large m and n (i.e. if m and n are greater than about 8). Thus for our samples with large sample sizes, we can use the normal approximation. If we test the hypothesis H_0: $F_X(z) = F_Y(z)$ for all z against the alternative hypothesis H_1: $F_X(z) \neq F_Y(z)$ for one z the test would be the following

$$\text{Reject } H_0 \text{ if } |T_X| \geq k \tag{7.1}$$

where k is determined by fixing the size of the test and using the asymptotic normal distribution of T_X. If the size of the test is fixed at 0.05 the critical value k equals 1.96. The mean and variance of T_X are (see Mood, Graybill and Boes (1974))

$$E(T_X) = \frac{m(m+n+1)}{2} \tag{7.2}$$

$$Var(T_X) = \frac{mn(m+n+1)}{12} \tag{7.3}$$

If T_X is large the values of X tend to be larger than the values of Y and this suggests that $F_X(.) \leq F_Y(.)$. On the other hand a small value of T_X suggests that $F_X(.) \geq F_Y(.)$.

- Kolmogorov-Smirnov test

The second test for H_0: $F_X(z) = F_Y(z)$ for all z is the Kolmogorov Smirnov test which makes use of the fact that the sample c.d.f. can be used to estimate the population c.d.f. The sample c.d.f. is defined by

$$F_{Xm}(x) = \frac{1}{m}.(\text{number of } X_i \text{ less than or equal to } x) \tag{7.4}$$

It can be shown that for fixed x $F_{Xm}(x)$ is an unbiased and mean-squared error consistent estimator of $F_X(x)$, regardless of the form of $F_X(x)$. The Glivenko-Cantelli-theorem

$$P\left(\sup_{-\infty<z<\infty} \left| F_{Xm}(z) - F_X(z) \right| \underset{m\to\infty}{\to} 0 \right) = 1 \tag{7.5}$$

states that the estimating function $F_{Xm}(x)$ of the c.d.f. $F_X(x)$ converges to $F_X(x)$ uniformly for all x with probability one.

So in the case that H_0 is true we have two independent estimators of the common population c.d.f., one using the c.d.f. of the X's and one using the sample c.d.f. of the Y's. The Kolmogorov-Smirnov test uses the closeness of the two sample c.d.f.'s to each other as a test criterion. The test statistic is

$$D_{m,n} = \sup_{-\infty<z<\infty} \mid F_{Xm}(z)-F_{Yn}(z) \mid \tag{7.6}$$

Kolmogorov showed that the statistic $D_m = \sup_{-\infty<z<\infty} \mid F_{Xm}(z)-F_X(z) \mid$ is asymptotically distribution free if H_0 holds. The hypothesis is rejected if D_m is large. The critical value of the asymptotic test statistic D_m can be approximated by $1.3581/\sqrt{m}$. Smirnov showed that $D_{m,n} / \sqrt{1/m+1/n}$ has the same limiting distribution as $D_m/\sqrt{m}$. See Kendall and Stuart (1979), volume 2, pages 476-487.

- Two-sample Pearson χ^2 test

The strength of the two aforementioned statistics is that they are distribution free. The χ^2 test statistic we describe below makes use of the multinomial distribution. The reason for presenting this test is that we will also use χ^2 tests in the next section. Let us partition the real line into J mutually disjoint sets $A_1, \ldots, A_J$. Define

$$p_{j1} = P(X \in A_j),\ j=1,\ldots,\ J \tag{7.7}$$

$$p_{j2} = P(Y \in A_j),\ j=1,\ldots,\ J \tag{7.8}$$

If $F_X(z) = F_Y(z)$ for all z then $p_{j1}=p_{j2}$, $j=1,\ldots, J$. Thus the hypothesis $F_X(z) = F_Y(z)$ for all z can be replaced by the hypothesis

$$H_0\colon p_{j1} = p_{j2}\ ,\ j=1,\ldots,\ J \tag{7.9}$$

In fact we then test the equality of two independent multinomial distributions.

Let us consider two independent multinomial distributions with parameters n_ℓ, $p_{1\ell}, \ldots, p_{J\ell}$, $\ell=1, 2$ respectively. Let $X_{j\ell}$, $j=1,\ldots, J$, $\ell=1,2$ represent the corresponding frequencies. It follows from the multivariate form of the Central Limit Theorem that the observed numbers $X_{j\ell}$ in the J groups will tend, for $n_\ell\to\infty$, to a multivariate normal distribution. Making use of the assumption of independence between X_{j1} and X_{j2} the quadratic form

$$\sum_{j=1}^{J} \frac{(X_{j\ell}-n_\ell p_{j\ell})^2}{n_\ell p_{j\ell}} \quad \ell=1,\ 2 \tag{7.10}$$

is distributed approximately as $\chi^2(k-1)$. The random variable

$$\sum_{\ell=1}^{2} \sum_{j=1}^{J} \frac{(X_{j\ell}-n_\ell p_{j\ell})^2}{n_\ell p_{j\ell}} \tag{7.11}$$

which is the sum of two stochastically independent random variables is approximately $\chi^2(2k-2)$.

Since the $p_{j\ell}$'s are unspecified, we need estimates of these parameters. It can be shown (Kendall and Stuart (1979), volume 2, pages 448-490) that the random variable

$$\sum_{\ell=1}^{2} \sum_{j=1}^{J} \frac{(X_{j\ell}-n_\ell \hat{p}_j)^2}{n_\ell \hat{p}_j} \tag{7.12}$$

has an approximate χ^2-distribution with 2k-2-(k-1)=k-1 degrees of freedom, where $\hat{p}_j$ is the maximum likelihood estimator of $p_{j1}=p_{j2}$ based upon the frequencies $X_{j\ell}$

$$\hat{p}_j = \frac{X_{j1}+X_{j2}}{n_1+n_2} \quad j=1,\ldots,\ J \tag{7.13}$$

Hypothesis H_0 is rejected when the computed value of the random variable in equation (7.13) is at least as great as some critical value.

- Test results

The two samples we use, OSA and SEP, are supposed to be drawn from the same subset of the Dutch population. In this section we want to find out whether this hypothesis can be maintained. The means and standard deviations of some important variables may shed some light on this problem. See also Tables 2.2 and 2.5 in Chapter 2.

Table 7.1 Sample means and standard deviations

	actual hours	preferred hours	pred. wage rate
single males OSA	30.75(17.83)	28.08(15.76)	12.94(1.63)
single males SEP	26.47(19.46)	32.85(14.91)	12.48(2.37)
married males OSA	39.70(11.70)	36.49(10.69)	13.61(1.64)
married males SEP	39.41(12.54)	38.23(10.39)	13.60(3.03)
single females OSA	23.63(19.61)	21.58(17.08)	11.31(1.75)
single females SEP	13.58(18.35)	17.04(17.97)	11.19(2.85)
married females OSA	10.64(15.44)	9.60(13.65)	11.13(1.52)
married females SEP	7.31(12.98)	10.00(13.32)	11.17(1.80)

We have performed the two-sample tests only on the two hours variables, actual and preferred, seperately. Of course it could very well be that the OSA and the SEP sample do not differ with respect to the hours variables but that they do differ with respect to one or more of the exogenous variables in our models. Only in the case of some kind of selection on the endogenous variables sample selection bias becomes a problem. Selection on the exogenous variables doesn't lead to biased estimates. In Table 7.2 the test results are presented. In some cases the various tests lead to conflicting conclusions. For all tests the null hypothesis is the equality of two samples. But the "implicit null" differs. In fact by the Kolmogorov-Smirnov is tested how well two given sets of observations fit each others sample c.d.f.. By the Pearson test is tested whether two independent multinomially distributed sets of observations are equal. And by the Wilcoxon we test whether the values of a variable in one sample tend to be larger than the values of that same variable in the other sample.

For single males the null hypothesis of equality is not rejected in most cases. This is in accordance with the conclusions that could be drawn if we look at the means in Table 7.1, which differ, but not significantly. For married males the sample means of actual hours are almost the same. Also according to the Wilcoxon test the null cannot be rejected. Both the Kolmogorov-Smirnov and the Pearson test reject the null. Rejection by the Pearson test occurs because the assumption of equality between the multinomial distributions is not true. The Kolmogorov-Smirnov tests indicate that the sample c.d.f.'s differ, although the sample means are close and although the values of the hours variable in one sample are not systematically lower or higher than in the other sample (see Wilcoxon test statistic). For equality between preferred hours in the two samples, it is the Kolmogorov-Smirnov test that does not reject and the other two tests that do reject the null hypothesis. Both for single and married females all tests point in the direction of inequality between the two samples.

Table 7.2 Two-sample tests, OSA against SEP[a)]

Complete samples, actual hours

	single males	married males	single females	married females	critical value
Wilcoxon	1.78	0.35	5.51^{*}	3.99^{*}	1.96
Kolm.-Smirnov	1.08	2.27^{*}	2.73^{*}	2.30^{*}	1.36
Pearson	22.60^{5}	55.19^{*4}	44.33^{*5}	43.88^{*3}	21.0-26.3[b)]

Samples without recipients of benefits, actual hours

	single males	married males	single females	married females	critical value
Wilcoxon	0.95	0.63	4.61^{*}	4.08^{*}	1.96
Kolm.-Smirnov	0.80	2.28^{*}	2.23^{*}	2.38^{*}	1.36
Pearson	25.29^{5}	56.75^{*4}	38.66^{*4}	46.30^{*3}	21.0-26.3[b)]

Complete samples, preferred hours

	single males	married males	single females	married females	critical value
Wilcoxon	-2.22^{*}	-3.67^{*}	2.78^{*}	-2.10^{*}	1.96
Kolm.-Smirnov	1.10	0.99	1.47^{*}	2.02^{*}	1.36
Pearson	23.84^{*2}	35.53^{*4}	23.45^{*3}	61.54^{*1}	21.0-26.3[b)]

Samples without recipients of benefits, preferred hours

	single males	married males	single females	married females	critical value
Wilcoxon	-0.11	-2.39^{*}	-3.38^{*}	-1.66	1.96
Kolm.-Smirnov	0.66	0.90	1.97^{*}	1.84^{*}	1.36
Pearson	12.37^{1}	34.81^{*4}	33.60^{*2}	60.48^{*3}	21.0-26.3[b)]

a) A * denotes that the null is rejected at the 95% confidence interval.

b) The critical value corresponding with 1 is 21.0, 2: 22.4, 3: 23.7, 4: 25.0, 5: 26.3.

7.3 χ^2 diagnostic tests

In this section we use a χ^2 goodness-of-fit statistic to test the null hypothesis that a model is correctly specified. The models we use consist of parametric families of conditional distributions of response variables (i.e. hours of work) given regressor variables. The test statistic has been derived by Heckman (1984). It is based on the differences between the sample distribution and the hypothesized distribution, generated by the model. First, the region in which the response variable lies is partitioned into disjoint cells in a nonrandom way. (In Andrews (1988) several methods of constructing random cells are discussed.) Then a quadratic form is calculated, based on the difference between the observed number of outcomes in each cell and the expected number according to the model. If the model is correct, the observed differences are small and the test statistic converges to a χ^2 random variable. If the model is incorrect, the quadratic form diverges to infinity as the sample size increases.

The test statistic is an adjustment of Pearson's chi-square test which makes use of the multinomial ML estimator. When the ordinary ML estimator is used, the distribution of Pearson's test statistic is bounded between two χ^2 variables (see Kendall and Stuart (1979)). In the test statistic proposed by Heckman the generalized inverse of the asymptotic covariance matrix of the aforementioned quadratic form is used as a weighing matrix. Therefore, this statistic has necessarily an asymptotic χ^2-distribution under the null hypothesis.

The following exposition closely follows Heckman (1984). The test statistic is based on the difference between the conditional c.d.f. of hours of work according to the model, $F_h(h|x,\vartheta)$, and the sample c.d.f.. The x is a vector of regressors and ϑ is estimated by maximum likelihood. Then

$$\sqrt{K}(\hat{\vartheta}-\vartheta) \overset{A}{\sim} N(0, \Sigma_\vartheta) \tag{7.14}$$

where K = number of observations

$\hat{\vartheta}$ = ML estimator of ϑ

Σ_ϑ = covariance matrix of the parameter vector ϑ

The region in which h lies is partitioned into J intervals

$$C_j = [c_j, c_{j+1}) \tag{7.15}$$

An indicator function for observation k is defined by

$$d_{kj} = \begin{cases} 1 \text{ if } h_k \in C_j \\ 0 \text{ otherwise} \end{cases} \tag{7.16}$$

The conditional expectation of observing h_k in C_j is

$$E(d_{kj}|x_k,\vartheta) = F_h(c_{j+1}|x_k,\vartheta) - F_h(c_j|x_k,\vartheta) \tag{7.17}$$

Then

$$d_{kj} = E(d_{kj}|x_k,\vartheta) + \varepsilon_{kj}, \; k=1,\ldots,K, \; j=1,\ldots,J \tag{7.18}$$

Since we do not know the true value of ϑ, we use the ML estimator of ϑ, $\hat{\vartheta}$. Then

$$d_{kj} = E(d_{kj}|x_k,\hat{\vartheta}) + u_{kj} \tag{7.19}$$

where $\quad u_{kj} = E(d_{kj}|x_k,\vartheta) - E(d_{kj}|x_k,\hat{\vartheta}) + \varepsilon_{kj}$ (7.20)

From Heckman (1984) we know that

$$\frac{1}{\sqrt{K}} \sum_{k=1}^{K} (d_k - E(d_k|x_k,\hat{\vartheta})) \overset{A}{\sim} N(0,\Sigma) \tag{7.21}$$

where $\quad d_k$ is a Jx1 vector of d_{kj}'s

$E(d_k|x_k,\hat{\vartheta}))$ is a vector consisting of the elements $E(d_{kj}|x_{kj},\hat{\vartheta}))$

From (7.21) it follows that

$$G= \frac{1}{\sqrt{K}} \sum_{k=1}^{K} (d_k - E(d_k|x_k,\hat{\vartheta}))' \; \Sigma^{-1} \frac{1}{\sqrt{K}} \sum_{k=1}^{K} (d_k - E(d_k|x_k,\hat{\vartheta})) \quad (7.22)$$

with Σ^{-1} defined as the Moore-Penrose generalized inverse is asymptotically χ^2 distributed with J-1 degrees of freedom.

Replacing ϑ by $\hat{\vartheta}$ in Σ yields a statistic based on $\hat{\Sigma}$ that converges to the same asymptotic distribution as the statistic based on Σ. For computation of G we need to compute Σ^{-1}. Andrews (1988) derives a consistent estimator of Σ^{-1} which leads to a test statistic that is much simpler to compute. In this chapter we will use this estimator, although the small sample properties are not as good as the estimator proposed by Heckman (see Andrews (1988)).

Test results

Table 7.3 provides an overall evaluation of the results obtained in this thesis. The estimated wage elasticities presented are reduced form elasticities. For all models we have calculated the goodness-of-fit statistic described in the previous section. All specifications are rejected at the 5 percent significance level. A reason for this is that there are almost certainly specification errors in these types of models due to the static instead of dynamic specification. To be able to estimate dynamic labour supply models we need panel data. Only recently such panel data have become available for The Netherlands. Given that we had to focus on static models it is interesting to compare the values of the statistics with each other to get an idea how much a certain extension improves the model specification. In the calculation of the test statistic we have selected 8 hours intervals: [0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64] for all groups. In some cases we had to combine two or more intervals to prevent zero probabilities in some intervals. In Appendix 7A we present the intervals used.

For nested models the values of the log likelihood ratio are also presented.

Let us first concentrate on the comparison of the results of one specific model estimated on two different data sets. According to the test reults presented in Table 7.2 the hypothesis of equality between the OSA and the SEP is not rejected for single males: Both samples are supposed to be drawn from the same population. In that case the estimated wage elasticity should be stable across samples. For both samples we find low elasticities not differing significantly from 0. For married males the two-sample tests do not lead to a uniform conclusion. Some reject and some do not reject the null hypothesis. Therefore, we cannot draw strong conclusions about the stability of parameters across samples. Looking at Table 7.3 suggests that for the group of married males the estimated wage elasticities are very low, but for the SEP they appear to be slightly higher than for the OSA. For married and single females the null hypothesis of equality between the two samples is rejected. For females the SEP sample yields higher estimated wage elasticities than the OSA. Apparently, selection bias is a serious problem here. The samples are no random drawings from the same subset of the Dutch population.

If we compare the labour supply wage elasticities of the different groups it strikes that the estimated elasticities of males are very small and in almost all cases not significantly different from 0. Single females have estimated wage elasticities of around 0.5. Married females respond more than single females to wage differences: the estimated wage elasticities vary around 2 with quite large differences across models.

Finally let us examine the differences between models estimated on one data set for one type of individual. In the first two rows of each panel in Table 7.3 the results are presented corresponding with our standard model with actual hours (a) as the dependent variable (Chapter 2). In the model of row 1 recipients of benefits are not included (nb), in row 2 they are (b). For males the differences in the wage elasticities are small. The χ^2 specification test suggests that adding the recipients of benefits doesn't improve the fit of the model. For females in the OSA sample the wage elasticity falls if recipients of benefits are included. But here too the χ^2 test points in the direction of a better fit if recipients are excluded. The same conclusions hold for the third and fourth row which corresponds with the same standard model except that now preferred hours (p) is the endogenous variable (Chapter 3). Apparently behaviour of individuals receiving unemployment benefits have not been modelled in a satisfactory way. Row 5 in the lower panels presents the results of the rationed versions (r) of the model corresponding with row 4.

Adding preference formation (pf) reduces the estimated wage coefficients in all cases (Chapter 4). The likelihood ratio statistics support the statistical significance of preference formation as do the values of the χ^2 statistic. These same statistics in rows 7 and 8 reveal once more that the separate influence of habit formation (hf) is overwhelming relative to preference interdependence (pi). It also becomes clear that it is the inclusion of habit formation that lowers the estimated wage elasticity. Measuring the influence of preference formation through a factor analysis model (row 9, fpf) yields results that are comparable to those presented in row 6.

Row 10 shows the OSA results corresponding with a model in which preferred hours and wages are simultaneously modelled, while recipients of benefits are excluded (p,w,nb) (Chapter 5). The value of the estimated wage elasticity hardly changes for males. For married females the elasticity has increased with a much lower standard error. Extending this model with the modelling of job choice leads to the results shown in rows 11 and 12. Two specifications were tried out for the production function of the amenity of a job, namely Z_2=hq (row11, p,w,hq,nb) and Z_2=q (row 12, p,w,q,nb). The elasticities corresponding with the Z_2=q specification have increased compared to the standard model. For married males and females incorporating job choice improves the fit significantly. For the other specification, Z_2=hq, the elasticities have remained the same, except for single females.

In row 13 results of the discrete choice model of actual hours and wages are given (Chapter 6). For the sake of simplicity, hours are assumed to be linear in wages and no recipients of benefits are included (a,lw,nb). Using a linear specification of hours in wages instead of a quadratic one, appears to increase the estimated wage elasticity considerably, especially for females. Apparently the quadratic wage effect is much more significant for females than for males. This is also borne out by the difference in the χ^2 statistics. Finally in row 14 hours restrictions are incorporated in the aforementioned model (a,lw,hr,nb). This reduces the wage elasticities for females, but they remain larger in size than in other, quadratic, specifications. Incorporating hours restrictions significantly improves the fit.

Table 7.3 Comparison of labour supply wage elasticities

single males	OSA			SEP		
	elas.(st.err.)	x^2	log Lik.	elas.(st.err.)	x^2	log Lik.
1) a,nb	0.04(0.16)	94.0	-	0.20(0.21)	93.1	-
2) a,b	0.08(0.16)	124.1	-	0.20(0.21)	141.5	-
3) p,nb	0.09(0.12)	87.4	-	-0.04(0.20)	91.9	-
4) p,b	0.09(0.12)	56.8	-400.3	0.03(0.20)	122.5	-518.8
6) p,b,pf	-0.03(0.13)	48.7	-388.9	-0.01(0.22)	126.0	-493.0
7) p,b,hf	-0.03(0.10)	48.7	-388.9	-0.01(0.22)	126.0	-493.0
8) p,b,pi	0.09(0.12)	56.8	-400.3	0.03(0.20)	122.5	-518.8
9) p,b,fpf	-	-	-	-0.03(0.15)	125.1	-492.1
10)p,w,nb	0.00(0.00)	76.2	-	-	-	-
11)p,w,hq,nb	0.00(0.00)	72.8	-	-	-	-
12)p,w,q,nb	0.10(0.13)	79.9	-	-	-	-
13)a,lw,nb	0.01(0.19)	52.8	-	0.00(0.00)	88.6	-
14)a,lw,hr,nb	0.01(0.01)	46.4	-	0.00(0.00)	67.1	-

married males	OSA			SEP		
	elas.(st.err.)	x^2	log Lik.	elas.(st.err.)	x^2	log Lik.
1) a,nb	-0.00(0.06)	311.2	-	0.14(0.04)	535.2	-
2) a,b	0.02(0.06)	337.2	-	0.11(0.03)	671.7	-
3) p,nb	-0.02(0.06)	427.6	-	0.08(0.04)	482.5	-
4) p,b	-0.01(0.06)	501.5	-	0.01(0.00)	538.1	-
5) p,b,r	0.02(0.03)	502.8	-4386.7	0.03(0.08)	646.7	-5220.4
6) p,b,r,pf	-0.00(0.01)	498.9	-4114.5	0.02(0.04)	679.3	-4735.0
7) p,b,r,hf	-0.00(0.01)	498.6	-4115.3	0.02(0.04)	679.3	-4735.0
8) p,b,r,pi	0.01(0.03)	502.5	-4380.6	-0.03(0.06)	632.6	-5196.4
9) p,b,r,fpf	-	-	-	-0.03(0.06)	592.6	-4722.6
10)p,w,nb	-0.02(0.02)	431.5	-	-	-	-
11)p,w,hq,nb	-0.02(0.03)	449.4	-	-	-	-
12)p,w,q,nb	0.00(0.00)	394.1	-	-	-	-
13)a,lw,nb	0.00(0.00)	388.8	-	0.03(0.02)	216.4	-
14)a,lw,hr,nb	0.00(0.00)	52.5	-	0.00(0.00)	141.6	-

Table 7.3 Comparison of labour supply wage elasticities, continued

single females	OSA			SEP		
	elas.(st.err.)	x^2	log Lik.	elas.(st.err.)	x^2	log Lik.
1) a,nb	0.43(0.24)	93.8	-	0.91(0.31)	88.5	-
2) a,b	0.23(0.08)	129.4	-	0.94(0.50)	174.4	-
3) p,nb	0.34(0.21)	91.2	-	0.82(0.31)	77.0	-
4) p,b	0.23(0.08)	141.2	-568.1	0.89(0.50)	128.2	-409.9
6) p,b,pf	0.11(0.11)	122.5	-538.1	0.11(0.24)	148.9	-345.4
7) p,b,hf	0.05(0.06)	121.7	-541.4	0.11(0.24)	148.9	-345.4
8) p,b,pi	0.22(0.09)	124.3	-546.9	0.65(0.42)	130.5	-404.6
9) p,b,fpf	-	-	-	0.09(0.22)	149.9	-344.2
10)p,w,nb	0.51(0.32)	76.2	-	-	-	-
11)p,w,hq,nb	0.01(0.11)	74.0	-	-	-	-
12)p,w,q,nb	0.51(0.34)	82.2	-	-	-	-
13)a,lw,nb	0.43(0.04)	74.8	-	3.48(3.15)	71.0	-
14)a,lw,hr,nb	0.27(0.17)	41.6	-	2.76(2.32)	60.3	-

married females	OSA			SEP		
	elas.(st.err.)	x^2	log Lik.	elas.(st.err.)	x^2	log Lik.
1) a,nb	2.04(3.92)	508.9	-	4.72(4.22)	583.7	-
2) a,b	1.83(1.28)	522.3	-	5.02(5.15)	628.2	-
3) p,nb	2.00(2.66)	413.2	-	4.95(4.81)	276.9	-
4) p,b	1.83(1.38)	441.3	-	5.05(2.89)	308.1	-
5) p,b,r	1.74(5.28)	440.4	-4386.7	2.17(4.21)	406.2	-5220.4
6) p,b,r,pf	0.39(0.51)	480.3	-4114.5	0.69(0.63)	224.6	-4735.0
7) p,b,r,hf	0.33(0.53)	479.9	-4115.3	0.69(0.63)	224.6	-4735.0
8) p,b,r,pi	0.53(0.64)	451.8	-4380.6	1.86(3.09)	465.6	-5196.4
9) p,b,r,fpf	-	-	-	0.48(0.51)	230.7	-4722.6
10)p,w,nb	1.64(0.34)	543.5	-	-	-	-
11)p,w,hq,nb	1.64(0.08)	541.4	-	-	-	-
12)p,w,q,nb	2.37(1.59)	531.5	-	-	-	-
13)a,lw,nb	3.36(1.74)	187.4	-	7.09(7.08)	137.5	-
14)a,lw,hr,nb	1.85(0.58)	118.6	-	1.84(0.34)	215.5	-

In Table 7.3 the following abbreviations are used

a,nb = actual hours, no recipients of benefits included
a,b = actual hours, recipients of benefits included
p,nb = preferred hours, no recipients of benefits included
p,b = preferred hours, recipients of benefits included
p,b,pf = preferred hours, recipients of benefits included, preference formation
p,b,hf = preferred hours, recipients of benefits included, habit formation
p,b,pi = preferred hours, recipients of benefits included, preference interdependence
p,b,fpf = preferred hours, recipients of benefits included, factor analysis model on preference formation
p,w,nb = simultaneous preferred hours, wage model, no recipients of benefits included
p,w,hq,nb = simultaneous preferred hours, wage, job choice model (specification Z_2=hq), no recipients of benefits included
p,w,q,nb = simultaneous preferred hours, wage, job choice model (specification Z_2=q), no recipients of benefits included
a,lw,b = simultaneous actual hours, linear wage model, no recipients of benefits included
a,lw,hr,b = simultaneous actual hours, linear wage model with hours restrictions, no recipients of benefits included

7.4 Concluding remarks

The two-sample tests discussed in Section 7.2 point in the direction of equality between the OSA and the SEP sample only for single males and to a lesser extent for married males. With respect to single and married females the tests unambiguously reject the hypothesis that the OSA and the SEP samples have been drawn from the same subset of the Dutch population.

In Section 7.3 an overall evaluation of the results obtained in this thesis has been provided. Single and married males appear to have very low wage elasticities, close to 0. Female labour supply wage elasticities are estimated to be around 0.5 for singles and around 2 for married females, varying from model to model. Extending the standard model with preference formation or hours restrictions improves the fit significantly. Both lead to lower wage elasticities. Modelling hours, wages and job choice simultaneously leads to less obvious conclusions. Much depends on the exact specification of job choice. Little variation in the data is probably a major problem. Overall the direction in which the results change as a consequence of a certain extension of the model is fairly stable across samples. Unfortunately, the estimated values of the elasticities corresponding to one particular version of the model vary a lot over the samples.

Appendix 7A Partitioning of the endogenous variable

actual hours worked, recipients of benefits not included

single males, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married males, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single males, SEP	[0,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,64]
married males, SEP	[0,33], (33,41], (41,49], (49,57], (57,64]
married females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,64]

actual hours worked, recipients of benefits included

single males, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married males, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single males, SEP	[0,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,64]
married males, SEP	[0,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,64]

<u>preferred hours worked, recipients of benefits not included</u>

single males, OSA	[0,25], (25,33], (33,41], (41,64]
single females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,64]
married males, OSA	[0,33], (33,41], (41,49], (49,57], (57,64]
married females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,64]
single males, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married males, SEP	[0,33], (33,41], (41,49], (49,57], (57,64]
married females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,64]

<u>preferred hours worked, recipients of benefits included</u>

single males, OSA	[0,1], (1,25], (25,33], (33,41], (41,64]
single females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,64]
married males, OSA	[0,17], (17,25], (25,33], (33,64]
married females, OSA	[0,1], (1,9], (9,17], (17,25], (25,33], (33,64]
single males, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
single females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married males, SEP	[0,17], (17,25], (25,33], (33,41], (41,49], (49,57], (57,64]
married females, SEP	[0,1], (1,9], (9,17], (17,25], (25,33], (33,41], (41,64]

8 Conclusion

The main point of emphasis in this thesis is the modelling of a neoclassical labour supply model for one-adult and two-adult households. The number of hours an individual prefers to work is assumed to follow from maximization of a utility function subject to a budget constraint. Throughout this thesis we have adopted the Hausman-Ruud specification of the utility function which leads to a labour supply equation quadratic in wages with the exception of Chapter 6 where we have opted for the linear Hausman specification. This was dictated by the complexity of the model in Chapter 6. The models developed in Chapters 2 through 6 have been estimated for four different groups of individuals: single males, single females, married males and married females. We distinguished between these four groups because we expected labour supply behaviour of these groups to be different. The estimation results confirm these expectations. Where possible, the models have been estimated on two different data sets, namely the SEP and the OSA Survey. This opened up the possibility of testing the stability of the parameter estimates across samples.

Estimation of the standard model in Chapter 2 yields male labour supply wage elasticities close to 0. Female elasticities vary from 0.01 for single females in the OSA to 7.09 for married females in the SEP. The simulated hours distributions fit the actual hours distributions poorly. This is commonly encountered in the literature on Tobit models. A possible reason for this could be that all nonworkers are implicitly assumed not to want to work. While in reality, however, there are unemployed workers who want to work but cannot find a job. Blundell et al. (1987) (following Cragg (1971)) developed a double hurdle model in which a standard Tobit model was modified to allow for this. In the participation decision demand side factors such as the unemployment rate and the vacancy rate play a large role. In Chapter 6 we allow for similar demand side restictions in a somewhat different way.

The largest part of the thesis however, is devoted to modelling the supply side of the labour market. In Chapter 3 a modification of our standard model is formulated taking explicit account of nonconvexities in the budget set, related to the Dutch social security and welfare system. This provides us with a more realistic model. Chapter 3 shows that adapting the standard model by also modelling behaviour of individuals receiving a benefit hardly changes the estimation results.

As in most econometric studies of labour supply we assume that the number of hours an individual prefers to work depends on a single wage rate independent of the number of working hours. If the income tax system is progressive (as is the case in The Netherlands) this assumption is not correct. Modelling the progressive tax system, however, leads to nonlinear budget sets. In Burtless and Hausman (1978) an algorithm is presented to estimate such models. An application of this on models of labour supply can be found in Van Soest, Woittiez and Kapteyn (1989). We need not use this algorithm when we exploit answers to a specific question in both our surveys about how many hours an individual would like to work (Chapters 3 through 5). These answers are referred to as the respondent's preferred hours. While asking the respondent for these preferred hours it is pointed out that the average after tax wage rate would remain constant. In other words, the respondent is offered a linear budget constraint. We had expected that the wage elasticities would be lower in the actual hours version of the supply equation than in the preferred hours version, since actual hours are partly determined by institutional constraints. In most cases we found however that the wage elasticities were slightly (not significantly) lower in the preferred hours version (Chapter 3).

When predicting labour supply over a longer period it seems hardly reasonable to assume that preferences remain constant. Chapter 4 therefore allows for habit formation (present tastes depend on own past behaviour) and interdependent preferences (present tastes depend on past behaviour of others). An important determinant of preferences is what happens in the social reference group of a household. In the first part of Chapter 4 social reference group variables have been constructed on the basis of a set of assumptions. The significance of reference group influences for the labour supply of households is confirmed by the data. In the second part of Chapter 4 direct information in the SEP data on reference groups has been used to construct a latent variables model of the influence of reference groups on labour supply. This latent variables model strengthens our earlier conclusions about the determinants of shifts in preferences. The results indicate that direct information improves the accuracy of the estimates and the statistical significance of preference formation. In all estimated versions the influence of habit formation relative to preference interdependence is overwhelming. One of the most interesting conclusions of this chapter is that the inclusion of preference formation leads to a decrease in the short run labour supply wage elasticities in all cases.

In general, a labour market transaction can be viewed as a tied sale in which the worker sells labour services and buys the attributes of his job. Hence the worker can maximize utility defined over job attributes by choosing the appropriate type of job and employer. Very few studies, however, incorporate nonpecuniary job characteristics into a utility maximization model. In Chapter 5 we develop such a structural model of job choice, labour supply and wages. The actual wage paid in the market results from two transactions, one for labour services and worker characteristics and another for job characteristics. The amount of money which makes an individual indifferent between work with good working conditions and work with bad working conditions is called his reservation wage differential. For married males the estimated reservation wage differential for clean work is about 6.6% of his income. In models in which job choice is not modelled the differential is much smaller and insignificant (see Chapter 5). In much empirical research on wage differentials the estimated differential is small and statistically not significant. This could possibly be due to the simultaneity bias as these studies do not take account of the endogeneity of job choice. Our results are not uniformly satisfactory, especially not for other groups than married males. Note that the number of individuals in these groups who are working in jobs with bad working conditions is very small.

Two possible extensions of the model in Chapter 5 are worth mentioning. First, note that the model is static and describes the equilibrium state of a dynamic process. An interesting subject of future research could be the development of a model of job choice in a search theoretic framework (see for example Burdett et al. (1984)). Second, an important problem in job choice models is the measurement of job characteristics. Our data set contains a large number of variables on working conditions. A factor analysis model could be used to estimate the real but unknown job amenity. This is left for future reasearch.

Many workers do not have complete flexibility in choosing their working hours. One can argue that the effective choice is between working a standard full time week, 20 hours per week or not working at all, whereby 20-hour jobs are relatively scarce. Individuals who prefer to work 20 hours per week, but who can only find a 40-hour job, may decide not to work. If that is the case unemployment is the result of hours restrictions. The labour supply decision can be described as the choice between a set of distinct points representing job opportunities. Each alternative in this choice set is offered with an unknown probability. In Chapter 6 we present and estimate a model that allows for the simultaneous estimation of these probabilities and the labour supply parameters. This model allows us to capture the sample distribution of working hours very well, for males as well as for females, both single or married. The rare occurrence of individuals working only a few hours per week is explained in this model by hours constraints. One should note that this explanation is a plausible interpretation, among many others. Other interpretations could be the existence of fixed costs on the supply side of the labour market, unobserved preference variation, or unobserved wage variation.

Identification of the model in Chapter 6 relies on several assumptions. We could get rid of these identifying assumptions if we could use independent information on the supply side of the labour market, about preferred hours, and on the demand side, about job offers. In Renes (1989) a model on vacancies has been developed and estimated in which the existence of various types of vacancies (such as part-time) is explained. Integration of this model and the model developed in Chapter 6 of this thesis seems very promising.

Finally in Chapter 7 we present an overall evaluation of the empirical results of this thesis. In all estimated models single and married males appear to have wage elasticities close to 0, while female elasticities vary mainly from about 0.5 to about 5. Our estimated values are in line with other estimates of Dutch elasticities as reported in a survey article by Theeuwes (1988). Male elasticities are reported to vary from -0.25 to 0.27, while female elasticities show a much larger range form 0.20 to 3.23. The simultaneous model of job choice, labour supply and wages leads to few obvious conclusions. Goodness of fit tests show that extending the standard model with preference formation or hours restrictions improves the fit significantly. Both extensions lead to lower wage elasticities. Given the statistical significance of these extensions it seems that a realistic model of labour supply cannot do without them.

The differences in estimates for the different samples, supposedly drawn from the same subset of the Dutch population, are disconcerting. Apparently we have a selection problem at hand. This demonstrates how important sampling is. Moreover it warns us to be careful in drawing strong conclusions on the basis of one set of estimates.

Several issues for future research have been mentioned already. Modelling job choice in a search theoretic framework is one of them. Using a factor analysis model to obtain estimates of the real but unknown amenity of a job is another. Also integrating a model on vacancies into the hours restrictions model seems promising. The availability of panel data calls for dynamic specifications of labour supply models. Finally, an integration of all extensions, if feasible, could lead to a more realistic model of labour supply.

References

Abowd, J.M. and O. Ashenfelter (1981), "Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Diffferentials", in: S. Rosen (ed.), Studies in Labor Markets, NBER.

Alessie, R. and A. Kapteyn (1985), "Habit Formation and Interdependent Preferences in the Almost Ideal Demand System," Working Paper, Tilburg University.

Altonji, J.G. and C.H. Paxson (1985), "Job Characteristics and Hours of Work", Working Paper, Princeton University.

Andrews, D. (1988), "Chi-square Diagnostic Tests for Econometric Models", Journal of Econometrics, 37, 135-156

Antos, J.R. and S. Rosen (1975), "Discrimination in the Market for Public School Teachers", Journal of Econometrics 3, 123-150.

Ashenfelter, O. (1980), "Unemployment as Disequilibrium in a Model of Aggregate Labor Supply", Econometrica, 48(3), 547-564.

Atrostic, B.K. (1982), "The Demand for Leisure and Nonpecuniary Job Characteristics", American Economic Review , 428-440.

Barnett, W.A. (1977), "Pollak and Wachter on the Household Production Function Approach", Journal of Political Economy, 85, 1073-1081.

Barron, J.M. and M.A. Loewenstein (1986), "On Imperfect Evaluation and Earnings Differentials", Economic Inquiry, 595-614.

Barten, A.P. (1969), "Maximum Likelihood Estimation of a Complete System of Demand Equations", European Economic Review, 7-73.

Barzel, Y. (1973), "The Determination of Daily Hours and Wages", Quarterly Journal of Economics, 87, 220-238.

Becker, G.S. (1965), "A Theory of the Allocation of Time", Economics Journal, 75, 493-517.

Becker, G.S. (1974), "A Theory of Social Interactions", Journal of Political Economy, 82, 1063-1093.

Bekkering, J.M., Y.K. Grift, G. Ridder and J.J. Siegers (1987), "Taxation, Social Premiums and Labour Supply of Married Women in Holland", Stichting voor Economisch Onderzoek der Universiteit van Amsterdam.

Bjorn, P.A. and Q.H. Vuong (1984), "Simultaneous Equations Models for Dummy Endogenous Variables: a Game Theoretic Formulation with an Application to Labour Force Participation", Discussion Paper, California Institute of Technology.

Blackorby, C. and D. Donaldson (1988), "Adult-Equivalence Scales and the Economic Implementation of Interpersonal Comparisons of Well-Being, working paper, University of British Columbia.

Blanciforti, L. and R. Green (1983), "An Almost Ideal Demand System incorporating Habits: An Analysis of Expenditure on Food and aggregate Commodity Groups", The Review of Economics and Statistics, 65, 3, 511-516.

Blomquist, S.N. (1983), "The Effect of Income Taxation on the Labor Supply of Married Men in Sweden", Journal of Public Economics, 22, 169-197.

Blundell, R.W., J. Ham and C. Meghir (1987), "Unemployment and Female Labour Supply", Economic Journal, 97, 44-64.

Blundell, R. and I. Walker (1982), "Modelling the Joint Determination of Household Labour Supplies and Commodity Demands", The Economic Journal, 92, 351-364.

Brown, J.R. and S. Rosen (1982), "On the Estimation of Structural Hedonic Price Models", Econometrica, 50, 765-769.

Burdett, K.N., N.M. Kiefer, D.T. Mortenson and G.R. Neumann (1984), "Steady States as Natural Rates in a Dynamic Discrete Choice Model of Labor Supply, in: G.R. Neumann and N.C. Westergård-Nielsen (eds.), Studies in Contempory Economics, Berlin, Springer Verlag.

Burtless, G. and J.A. Hausman (1978), "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment", Journal of Political Economy, 86, 1103-1130.

Central Bureau of Statistics (1986), "Statistisch Zakboek 1986".

Cragg, J.G. (1971), "Some Statistical Models for Limited Dependent Variables with Applications to the Demand for Durable Goods", Econometrica, 39, 829-844.

Darrough, M., R. Pollak and T. Wales (1983), "Dynamic and Stochastic Structure: An Analysis of Three Time Series of Household Budget Studies", Review of Economics and Statistics, 65, 274-281.

Deaton, A. and J. Muellbauer (1980), "Economics and Consumer Behavior", Cambridge: Cambridge University Press.

Dickens, W. and S. Lundberg (1985), "Hours Restrictions and Labor Supply", NBER Working Paper No. 1638.

Duncan, G.J. (1976), "Earnings Functions and Nonpecuniary Benefits", The Journal of Human Resources, 11, 462-483.

Festinger, L. (1954), "A Theory of Social Comparison Processes", Human Relations, 7, 117-140.

Gaertner, W. (1974), "A Dynamic Model of Interdependent Consumer Behavior", Zeitschrift für Nationalökonomie, 34, 327-344.

Graham, J. W. and C. A. Green (1984), "Estimating the Parameters of a Household Production Function with Joint Products", Review of Economic Statistics, 66, 277-282.

Gronau, R. (1977), "Leisure, Home Production and Work-The Theory of the Allocation of Time revisited", Journal of Political Economy, 85, 1099-1123.

Gronau, R. (1980), "Home Production- A Forgotten Industry", Review of Economic Statistics, 62, 408-415.

Gronau, R. (1986), "Home Production- A Survey", in: Handbook of Labor Economics, eds. O. Ashenfelter and R. Layard, Amsterdam: North-Holland.

Hacken, P.T.M. ten, A. Kapteyn and I. Woittiez (1989), "Unemployment Benefits and the Labour Market, A Micro/Macor Approach", in: The Political Economy of Social Security, eds. B.A. Gustafson and N.A. Klevmarken, Amsterdam, North Holland

Hamermesh D.S. (1977), "Economic Aspects of Job Satisfaction", in: Essays in Labor Market Analysis, eds. O. Ashenfelter and W. Oates, New York: Wiley.

Hausman, J. (1980), "The Effect of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation", Journal of Public Economics, 14, 161-194.

Hausman, J. (1985), "The Econometrics of Nonlinear Budget Sets", Econometrica, 53, 1255-1282.

Hausman, J. and P. Ruud (1984), "Family Labor Supply with Taxes", American Economic Review, 74, 242-248.

Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error", Econometrica, 47, 153-161.

Heckman, J.J. (1984), "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated From Micro-data, Econometrica, 52, 1543-1547.

Kapteyn, A., P. Kooreman and A. van Soest (1989), "Quantity Rationing and Coherency in a Flexible Household Labor Supply Model", working paper, Tilburg University, forthcoming in The Review of Economics and Statistics.

Kapteyn, A., S.A. van de Geer, H. van de Stadt, T.J. Wansbeek (1985), "Interdependence of Preferences: An Econometric Analysis," Working Paper, Tilburg University.

Kapteyn, A. and I.Woittiez (1986), "Arbeidsaanbod en Voorkeursvorming", OSA werkdocument, nr. W23.

Kapteyn, A. and I. Woittiez (1989), "Preference Interdependence and Habit Formation in Family Labor Supply", in: Microeconometrics: Surveys and Applications, ed. by F. Laisney et al., Clevedon, Avon, Tieto Ltd., Chapter 10.

Kapteyn, A., I.Woittiez and P. ten Hacken (1989), "Household Labor Supply in The Netherlands in The Eighties and The Nineties", OSA werkdocument, nr. W 61.

Kendall, M.G. and A. Stuart (1979), "The Advanced Theory of Statistics, 2, New York, Charles Griffin and CO.

Killingsworth, M.R. (1984), "A Single Structural Model of Heterogeneous Preferences and Compensating Wage Differentials", Working Paper, Rutgers State University.

Killingsworth, M.R. and J.J. Heckman (1986), "Female Labour Supply: A Survey", in: Handbook of Labor Economics, eds. O. Ashenfelter and R. Layard, Amsterdam: North-Holland.

Kooreman, P. and A. Kapteyn (1985), "Estimation of a Game Theoretic Model of Household Labor Supply," Working Paper, Tilburg University.

Lancaster, K.J. (1966), "A new Approach to Consumer Theory", Journal of Political Economy, 74, 132-157.

Lucas, R.E.B. (1974), "The Distribution of Job Characteristics", The Review of Economics and Statistics, 530-540.

Maddala, G.S. (1983), "Limited Dependent and Qualitative Variables in Econometrics", Cambridge: Cambridge University Press.

Mann, H.B., and D.R. Whitney (1947)," On a Test of Whether One of Two Variables is Stochastically Larger Than the Other", Annals of Mathematical Statistics, 18, 50

Manser, M. and M. Brown (1980): "Marriage and Household Decision Making: a Bargaining Analysis", International Economic Review, 21, 31-44.

Marcus, M. and H. Minc (1964), "Matrix Theory and Matrix Inequalities", Boston: Prindle, Weber and Schmidt.

McCormick, (1986), "Employment Opportunities, Earnings, and the Journey to Work of Minority Workers in Great-Britain", The Economic Journal, 96, 375-397.

McElroy, M.B. and M.J. Horney (1981), "Nash-Bargained Household Decisions: Towards a Generalization of the Theory of Demand", International Economic Review, 22, 333-347.

Ministry of Social Affairs and Employment (1989), "Rapportage Arbeidsmarkt 1989".

Moffitt, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints", The Review of Economics and Statistics, 64, 510-515.

Moffitt, R. (1984): "The Estimation of a Joint Wage-Hours Labor Supply Model", Journal of Labor Economics, 2, 550-566.

Mood, A.M., F.A. Graybill and D.C. Boes (1974), "Introduction to the Theory of Statistics, Tokyo: McGraw-Hill

Muellbauer, J. (1986), "Habits, Rationality and Myopia in the Life-Cycle Consumption Function", Working Paper, Nuffield College, Oxford.

Neary, J.P. and K.W.S. Roberts (1980), "The Theory of Household Behaviour Under Rationing", European Economic Review, 13, 25-42.

Pencavel, J. (1986), "Labor Supply of Men: A Survey", in: Handbook of Labor Economics, eds. O. Ashenfelter and R. Layard, Amsterdam: North-Holland.

Phlips, L. (1984), "Applied Consumption Analysis", North-Holland, Amsterdam.

Pollak, R.A. (1970), "Habit Formation and Dynamic Demand Functions", Journal of Political Economy, 78, 745-763.

Pollak, R.A. (1976), "Interdependent Preferences", The American Economic Review, 66, 309-320.

Pollak, R.A. and M. L. Wachter (1975), "The Relevance of the Household Production Function and its Implications for the Allocation of Time", Journal of Political Economy, 83, 255-277.

Pollak, R.A. and T. Wales (1981), "Demographic Variables in Demand Analysis", Econometrica, 49, 1533-1551.

Renes, G. (1988), "Vacancy Durations", Research Memorandum, Leyden University, Leiden.

Roback, J. (1982), "Wages, Rents, and the Quality of Life", Journal of Political Economy, 90, 1257-1278.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", Journal of Political Economy, 82, 34-55.

Rosen, H.S. (1976), "Taxes in a Labor Supply Model with joint Wage-Hours Determination", Econometrica, 44, 485-507.

Rosen, S. (1979), "Wage-based Indexes of Urban Quality of Life", in: Current Issues in Urban Economics, eds. P. Mieszkowsky, M. Straszheim, Baltimore: Johns Hopkins University Press.

Schmidt, P. (1981), "Constraints on the Parameters in Simultaneous Tobit and Probit Models", in: Structural Analysis of Discrete Data: With Econometric Applications, ed. C.F. Manski and D. McFadden, Cambridge: The MIT Press.

Scotchmer, S. (1985), "Hedonic Prices and Cost/Benefit Analysis", Journal of Economic Theory, 37, 55-57.

Scott, R. (1972), "Avarice, Altruism and Second Party Preferences", The Quarterly Journal of Economics, 86, 1-18.

Van Soest, A., I. Woittiez and A. Kapteyn (1989), "Labour Supply, Income Taxes and Hours Restrictions in The Netherlands", CentER Discussion Paper, Tilburg Uiversity.

Van Soest, A., P. Kooreman and A. Kapteyn (1988), "Coherent specification of demand systems with corner solutions and endogenous regimes", Research Memorandum, Tilburg University.

Van de Stadt, H., A. Kapteyn and S.A. van de Geer (1985): "The Relativity of Utility: Evidence from Panel Data", The Review of Economics and Statistics, 67, 179-187.

Takayama, A. (1985), "Mathematical Economics", Cambridge: Cambridge University Press.

Theeuwes, J. (1988), "Arbeid and Belastingen", in: Belastingheffing en belastingvorming, Koninklijke Vereniging voor Staathuishoudkunde, Stenfert Kroese, Leiden, 111-143.

Tomes, N. (1983): "Income Distribution, Happiness and Satisfaction: A Direct Test of the Interdepent Preferences Model", Working Paper, University of Western Ontario.

Tummers, M. and I. Woittiez (1988), "A Simultaneous Wage and Labour Supply Model with Hours Restrictions", Research Memorandum, Tilburg University.

Viscusi, W.K. (1978), "Wealth Effects and Earnings Premiums for Job Hazards", Review of Economics and Statistics, 60, 408-416.

Wales, T.J. (1978), "Labour Supply and Commuting Time: An Empirical Study", Journal of Econometrics, 8, 215-226.

Witte, A.D., H.J. Sumka and H. Erekson (1979), "An Estimate of an Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets", Econometrica, 47, 1151-1173.

Woittiez, I.B. and A. Kapteyn (1987), "A Model of Job Choice, Labour Supply and Wages, Working Paper, Tilburg University.

Samenvatting

Een van de belangrijkste doelstellingen van politici in Nederland is het terugdringen van het hoge werkloosheidspercentage (11% in 1988). De omvang van de werkloosheid hangt zowel af van de vraag naar arbeid als van het aanbod van arbeid. Daarom is het niet alleen van belang maatregelen te nemen die een positieve invloed hebben op de werkgelegenheid, maar is het ook belangrijk de factoren te kennen die van invloed zijn op het arbeidsaanbod. In dit proefschrift trachten we het begrip daaromtrent te vergroten door middel van het construeren van gedetailleerde modellen van arbeidsaanbod.

Zo is bijvoorbeeld behalve de loonvoet ook de hoogte van de werkloosheids- en bijstandsuitkeringen een determinant van het arbeidsaanbod in onze modellen. Tevens worden verschillende aspecten geïncorporeerd die de preferenties ten aanzien van arbeidsaanbod beïnvloeden, zoals gezinssamenstelling, het verschijnsel dat onze huidige wensen afhangen van onze ervaringen in het verleden (gewoontevorming) en het verschijnsel dat onze wensen afhangen van wat we anderen zien doen. Ook wordt er bekeken in hoeverre niet-geldelijke karakteristieken van een baan van belang zijn voor arbeidsaanbod. Bovendien houden we rekening met het feit dat het aantal te werken uren per week vaak niet ter keuze van de werknemer is.

De data die we gebruiken hebben betrekking op Nederlandse huishoudens in 1985. Er wordt onderscheid gemaakt tussen verschillende groepen: eenpersoons huishoudens (alleenstaande mannen en alleenstaande vrouwen) en tweepersoons huishoudens. Door voor elk van deze groepen de verschillende modellen te schatten die in dit proefschrift ontwikkeld zijn, zouden de effecten van een verschuiving in de samenstelling van de beroepsbevolking geanalyseerd kunnen worden.

Na een inleidend hoofdstuk volgen vijf hoofdstukken waarin telkens een ander aspect van het arbeidsaanbod gedrag wordt belicht. In Hoofdstuk zeven worden de schattingsresultaten van de verschillende modellen naast elkaar gezet. De modellen worden onder andere vergeleken op basis van de geschatte loonelasticiteiten. Het geheel wordt gecompleteerd met een concluderend hoofdstuk. Opvallend is dat in bijna alle gevallen de loonelasticiteit van mannen klein is. De loonelasticiteiten van vrouwen

zijn groter, maar variëren sterk tussen de modellen. Vooral gewoontevorming blijkt een belangrijke invloed te hebben op het arbeidsaanbod van zowel mannen als vrouwen. Tenslotte heeft het rekening houden met het feit dat werknemers niet altijd het aantal uren kunnen werken dat ze zouden willen een daling van de elasticiteiten tot gevolg.

In Hoofdstuk twee staat het model beschreven dat als uitgangspunt dient voor de overige modellen. We beginnen met de analyse van het arbeidsaanbod gedrag van een huishouden dat bestaat uit één individu dat beschikbaar is voor de arbeidsmarkt (we zien dus af van een analyse van het gedrag van ouderen, zieken, studenten, militairen etc.). We veronderstellen dat het individu een optimale keuze moet maken tussen inkomen en vrije tijd. Die optimale keuze volgt uit de maximalisatie van een nutsfunctie, gegeven een lineaire budget restrictie. Uit de door ons gekozen specificatie van de nutsfunctie volgt een vraagvergelijking naar vrije tijd die kwadratisch is in het loon en lineair in de overige variabelen. Vervolgens wordt een model voor een tweepersoons huishouden geïntroduceerd, waarin de gezamenlijke arbeidsaanbod beslissing van man en vrouw wordt gemodelleerd.

Tevens wordt in Hoofdstuk twee een uitgebreide beschrijving van de data gegeven. Bij het schatten gebruiken we informatie van zowel werkende als niet werkende individuen, wat leidt tot een "Tobit" model. Omdat er voor niet werkenden geen loonvoet wordt waargenomen, schatten we op basis van gegevens van werkenden een loonvergelijking in termen van opleiding en leeftijd, waarbij we corrigeren voor "selection bias". Deze voorspelde lonen worden gebruikt in de modellen van de Hoofdstukken twee tot en met vier. In het vijfde en zesde hoofdstuk worden de parameters in de loon- en urenvergelijking simultaan geschat.

In Hoofdstuk drie worden eerst enkele complicaties van het Nederlandse sociale zekerheidssysteem geïntroduceerd. Als iemand, die werkloos is en een werkloosheidsuitkering ontvangt, een baan vindt, wordt hij/zij gekort op zijn/haar uitkering. Dit introduceert niet-convexiteiten in de budget restrictie. Indien de budget restrictie niet convex is, is het mogelijk dat er meerdere raakpunten zijn tussen de indifferentiecurven en de budget restrictie. Met behulp van nutsvergelijking wordt dan het punt corresponderend met het het hoogste nut gekozen en daarmee het optimale arbeidsaanbod bepaald.

Een andere aanpassing van de eenvoudige modellen uit het tweede hoofdstuk (en van veel modellen uit de empirische literatuur) is het gebruik van gewenste uren in plaats van feitelijke gewerkte uren. Omdat individuen vaak niet vrij het aantal uren kunnen kiezen dat ze willen werken, zullen in veel gevallen de gewenste en feitelijke uren niet overeenkomen. Door gewenste uren te gebruiken wordt een model verkregen waarin de vraagzijde van de arbeidsmarkt geen rol speelt. In Hoofdstuk zes wordt een model beschreven waarin vraagrestricties expliciet gemodelleerd zijn.

In Hoofdstuk vier ligt de nadruk op de invloed van het gedrag van anderen (interdependentie van preferenties) en van gewoontevorming op het arbeidsaanbod. Beide invloeden worden geoperationaliseerd door één van de parameters in het model afhankelijk te maken van het aantal uren dat het individu vorig jaar werkte en van het aantal uren dat mensen in zijn "sociale referentiegroep" vorig jaar werkten. In het eerste deel van Hoofdstuk vier worden de sociale referentiegroep variabelen geconstrueerd op basis van een aantal aannames. In het tweede deel wordt een factor analyse model gebruikt om de "werkelijke", maar niet geobserveerde, referentiegroep variabelen te schatten.

In Hoofdstuk vijf wordt de keuze van een baan, het aantal uren dat daarin gewerkt wordt en het loon dat men daarin verdient in een simultaan model beschreven. Nagegaan wordt in hoeverre niet-geldelijke karakteristieken van een baan van belang zijn voor arbeidsaanbod en loon ("wage differential theory").

Tenslotte wordt in Hoofdstuk zes een eerste stap gezet om de vraag naar arbeid te betrekken in arbeidsaanbod modellen. Dit is voornamelijk van belang als arbeidsvraag en arbeidsaanbod niet goed op elkaar aansluiten. Een voorbeeld daarvan is het tekort aan part-time banen. In dit model kiest het individu die baan (i.e. het urenaantal) uit de verzameling mogelijke banen die het hoogste nut oplevert.

In Hoofdstuk zeven worden de verschillende modellen uit de voorgaande hoofdstukken vergeleken. Zowel formele vergelijkingscriteria (statistische toetsen) als informele vergelijkingen op basis van gesimuleerde urenverdelingen en elasticiteiten worden besproken.

Conclusies uit de vergelijkingen van Hoofdstuk zeven zijn in het laatste hoofdstuk te vinden. De verschillen in schattingen tussen de twee steekproeven onderstrepen hoe belangrijk het is om goede data te verzamelen. Wat betreft de loonelasticiteiten kunnen we stellen dat die van mannen dichtbij nul liggen, terwijl die van vrouwen grofweg variëren tussen de 0.5 en 5 met uitschieters van 0.0 en 7.5. Dit betekent dat volgens onze modellen het arbeidsaanbod van vrouwen nogal toe zou nemen als gevolg van stijgende lonen. Statistische toetsen tonen aan dat het uitbreiden van het standaard model uit Hoofdstuk twee met interdependentie van preferenties en gewoontevorming (Hoofdstuk vier) en urenrestricties (Hoofdstuk zes) de "fit" significant verbetert. Beide uitbreidingen leiden tot lagere loonelasticiteiten dan anders het geval zou zijn. Het opnemen van urenrestricties in het model leidt tot een gesimuleerde urenverdeling die zeer goed de feitelijke verdeling benadert.

STELLINGEN

bij het proefschrift

MODELLING AND EMPIRICAL EVALUATION OF LABOUR SUPPLY BEHAVIOUR

van

ISOLDE WOITTIEZ

1. Het verwijt van Leontief (1971) dat econometrie slechts een poging is om de zwakheid van de verkrijgbare data te compenseren door het gebruik van steeds geavanceerdere technieken is onjuist. Het onderstreept niettemin het grote belang van goede dataverzameling.
   Zie Leontief, W. (1971), "Theoretical Assumptions and Nonobserved Facts", American Economic Review, 61, 1-7.

2. Het arbeidsaanbodgedrag van individuen wordt statistisch significant beïnvloed door gewoontevorming en door het gedrag van anderen in hun sociale omgeving. Het rekening houden met deze invloeden leidt tot lagere loonelasticiteiten dan anders het geval is.
   Zie Kapteyn, A., I. Woittiez en P. ten Hacken (1989)," Household Labor Supply in The Netherlands in the Eighties and the Nineties", OSA-werkdocument, W61 en Hoofdstuk 4 van dit proefschrift.

3. Indien in een nutsmaximalisatiemodel de prijzen afhangen van de beslissingsvariabelen, moeten de prijzen en de vraagfuncties simultaan geschat worden (zie Hoofdstuk 5 van dit proefschrift). In Rosen (1976) wordt dit ten onrechte niet gedaan.
   Zie Rosen, H.S. (1976)," Taxes in a Labor Supply Model with Joint Wage-Hours Determination", Econometrica, 44, 485-507.

4. Uit de schattingen van de preferentieparameters en de parameters die de beschikbaarheid van verschillende banen met bijbehorende urenaantallen aangeven, kan worden afgeleid dat de lage participatiegraad van getrouwde vrouwen in Nederland mede een gevolg is van het tekort aan part-time banen.
   Zie Hoofdstuk 6 van dit proefschrift.

5. Een nadeel van χ^2 goodness-of-fit testen is het informatieverlies dat onstaat als gevolg van het groeperen van observaties in klassen. Dit nadeel wordt gecompenseerd door het voordeel dat de asymptotische verdeling van de test-statistic bekend is.

6. Het vergelijken van loonelasticiteiten, zoals bijvoorbeeld in overzichtsartikelen van Pencavel (1986) en Killingsworth (1986) verliest veel aan betekenis als niet ook de bijbehorende standaardfouten gepresenteerd worden.
   Zie Pencavel, J. (1986)," Labor Supply of Men: A Survey", in: Handbook of Labor Economics, O.C. Ashenfelter and R. Layard (eds.), North-Holland, Amsterdam en Killingsworth, M.K. en J.J. Heckman (1986)," Female Labor Supply: A Survey", in: Handbook of Labor Economics, O.C. Ashenfelter and R. Layard (eds.), North-Holland, Amsterdam.

7. Kracht en zwakte van een niet-parametrische benadering van schattingsmethoden zijn beide gelegen in het zo min mogelijk opleggen van een structuur.

8. De theorie van de "compensating wage differentials" wordt ondersteund door de waarneming dat er nog steeds economen en econometristen werkzaam zijn als universitair docent.

9. Teneinde open te staan voor nieuwe ideeën strekt het tot aanbeveling om minstens één maal per jaar letterlijk en figuurlijk grote afstand te nemen van het werk.

10. De tolerantie ten aanzien van het gedrag van drugsverslaafden neemt recht evenredig af met de afstand tot hun verblijfplaats.

11. Er worden wel goede scheidsrechters (namelijk bonds-) ingezet bij derde klasse herenhockey en hoger en niet bij eerste klasse dameshockey en lager. Daaruit blijkt dat ruw spel beloond wordt met betere wedstrijdleiding.